

THIS REPORT HAS BEEN DELIMITED
AND CLEARED FOR PUBLIC RELEASE
UNDER DOD DIRECTIVE 5200.20 AND
NO RESTRICTIONS ARE IMPOSED UPON
ITS USE AND DISCLOSURE.

DISTRIBUTION STATEMENT A

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED.

AD903710

Naval Warfare Research Center
Information Science Laboratory

March 1972

Technical Report 1
NWRC TR-6

DYNAMIC ANALYSIS OF ASW EFFECTIVENESS — A QUEUING APPROACH

By: M. W. ZUMWALT, A. J. KOSAK, R. S. RATNER

Distribution limited to U.S. Gov't. agencies only;
Test and Evaluation; *gray 72* Other requests
Prepared for: *for this document must be referred to*

NAVAL ANALYSIS PROGRAMS (Code 462)
OFFICE OF NAVAL RESEARCH
ARLINGTON, VIRGINIA 22217

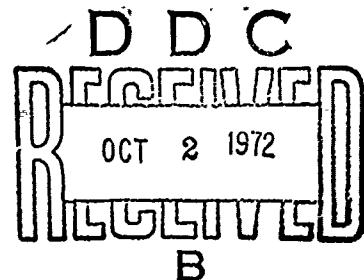
CONTRACT N00014-71-C-0417
Task No. NR 274-008-37

This is a report of progress on an analytic research task investigating possible application of queuing theory to ASW effectiveness analysis. As such, the concepts may be changed or qualified as the research continues. This report is prepared for distribution at this time to afford an opportunity for recipients to comment on the new techniques.

Reproduction in whole or in part is permitted for any purpose of the United States Government.



STANFORD RESEARCH INSTITUTE
Menlo Park, California 94025 • U.S.A.





STANFORD RESEARCH INSTITUTE
Menlo Park, California 94025 · U.S.A.

*Naval Warfare Research Center
Information Science Laboratory*

*Technical Report 1
NWRC TR-6*

March 1972

DYNAMIC ANALYSIS OF ASW EFFECTIVENESS — A QUEUING APPROACH

By: M. W. ZUMWALT, A. J. KOSAK, R. S. RATNER

Prepared for:

NAVAL ANALYSIS PROGRAMS (Code 462)
OFFICE OF NAVAL RESEARCH
ARLINGTON, VIRGINIA 22217

CONTRACT N00014-71-C-0417
Task No. NR 274-008-37

SRI Project 1319-1

Reproduction in whole or in part is permitted for any purpose of the United States Government.

Approved by:

LAWRENCE J. LOW, *Director*
Naval Warfare Research Center

DAVID R. BROWN, *Director*
Information Science Laboratory

ERNEST J. MOORE, *Executive Director*
Engineering Systems Division

ABSTRACT

This interim report demonstrates the feasibility of conducting a dynamic analysis of ASW effectiveness through the application of queuing theory. The ASW effectiveness of both the individual ASW unit and the aggregated ASW force are considered. Analytical models are formulated representing both levels of capability, and appropriate computational techniques are described and illustrated by selected examples. Areas requiring further research and development are defined, and recommendations for near term research are included which are in consonance with a continuing ASW evaluation program.

PREFACE

The research reported in this interim report was conducted as a portion of a continuing program directed toward the development of a queuing methodology for evaluating ASW system effectiveness. The project was sponsored by the Director, Naval Analysis Programs, of the Office of Naval Research. Mr. R. H. Dickman was the ONR Project Scientific Officer.

The research effort was performed jointly by the Naval Warfare Research Center, Mr. L. J. Low, Director, and the Information Sciences Laboratory, Mr. D. R. Brown, Director, of Stanford Research Institute. Mr. M. W. Zumwalt of NWRC was the project leader.

The following Stanford Research Institute personnel contributed directly to the accomplishment of the study:

Thomas L. Humphrey	Samuel Schechter
Andrew J. Korsak	Marvin W. Zumwalt
Robert S. Ratner	William H. Zwisler
Marilyn J. Rundberg	

In addition, Lieut. Thomas D. Kenneally, USN, and Lieut. Norman T. Saunders, USCG, from the U.S. Naval Postgraduate School, Monterey, California, provided significant analytical assistance in the implementation of the ASW force effectiveness model.

STUDY SUMMARY

Research Objective and Scope

The ultimate objective of the research reported on herein is to develop a queuing methodology for analyzing the effectiveness of ASW operations. The motivation for basing an ASW model on mathematical queuing is to account for congestion phenomena that may occur as several contacts are "serviced" by an ASW system. This service consists of the functions of detection, localization, attack, kill, classification, and command and control.

The present report does not contain the completed methodology. It does present the first necessary steps toward conceptualizing the technical aspects of modeling ASW operations in terms of queuing theory, and contains the calculation of a number of statistical parameters for the initial queuing models developed. The parameters identified are helpful for calculating ASW systems measures of effectiveness. One such parameter is the mean time to prosecute a contact from detection to attack and kill.

The report will be of interest primarily to those engaged in constructing models for ASW or other areas where congestion may significantly influence the assessment of effectiveness. It should also be useful to operational personnel by reason of its orientation.

Approach

The approach taken consists of two parts:

1. Analysis of an ASW unit with respect to detection, localization, attack, and kill functions

2. Analysis of an ASW force consisting of individual interacting units (such as the effects of detection capability).

To facilitate model formulation, only the four functions of detection, localization, attack, and kill have been considered. Classification and command and control were omitted for the time being but will be included later.

Findings

Conclusions

Specific elements of queuing theory have been found applicable to the modeling of congestion phenomena that might be encountered by individual surface ASW units and ASW force configurations. In the case of the ASW units, closed-form analytical results could not be obtained. However, suitable models have been developed using a synthesis of the theory of Markov processes and flow graph analysis techniques. Computer implementation of the models yielded probability distributions from which certain measures of effectiveness can be obtained. In the case of ASW force analysis, closed-form solutions have been obtained for direct computer calculation of particular measures of effectiveness, assuming simplified representation for the ASW units comprising the force. Examples of measures of effectiveness obtainable from the initial models are:

- Expected number of contacts lost while in the ASW system (over a given period of time)
- Probability of missing a true contact (failure to detect) for a given level of contact activity (false and true)
- Expected time to detection
- Expected time from detection to kill.

The current method of approach suggests that further detail in representing both individual ASW units and force configurations is possible without rendering the models computationally intractable. (The basic philosophy is that the Markov modeling approach for ASW units should be pursued as far as possible so long as computational tractability is maintained.) It appears, however, that whole ASW force models cannot be obtained simply by compounding detailed unit models, and that the two-pronged approach used so far will be most likely to succeed in further investigations. Information obtained on congestion behavior of ASW units will, of course, be of value in creating approximate models for their behavior in force models. As an example of this, the curves for distributions of saturated and unsaturated duration of the ASW unit considered thus far appear to be approximated by an exponential distribution. This result is used in the force model.

Potential Applications

Some of the applications of the completed methodology are as follows:

1. Comparison of congestion effects in an ASW unit or force with improved capability in the detection, classification, localization, attack, and kill functions. This could aid in the study of tradeoffs for concentrating effort in ASW operations to achieve maximal capability, as well as in decisions on where to focus development of improved hardware. For such purposes, modeling of the bottlenecks in an ASW system and how they interact in some degree of detail should prove valuable in assessing the effects of changes in system capabilities.
2. Evaluation of ASW force performance for alternative force mixes and force deployment geometries. The queuing methodology is unique in that it will provide an approach for assessing the dynamic changes in force performance due to changes in force mix and deployment geometry, that is, changes in force mix due to combat or operational attrition, and the changes in deployment geometry due to assignment of an ASW unit to prosecute a contact.

CONTENTS

ABSTRACT	iii
PREFACE	v
SUMMARY	S-1
LIST OF ILLUSTRATIONS	xi
LIST OF TABLES	xv
GLOSSARY OF TERMS	xviii
 I INTRODUCTION	 I-1
II GENERAL SITUATIONAL FORMULATIONS	II-1
A. ASW Functional Definitions	II-1
B. Functional Symbolic Algebra	II-4
III UNIT MODEL	III-1
A. Introduction	III-1
1. Description	III-1
2. Outline of Queuing Model	III-2
3. Measures of Effectiveness	III-5
4. Approach Selected for Model Analysis	III-7
B. Detailed Description of Single Element Model	III-8
1. Structure of the Queuing Situation in the Single Element Model	III-8
2. The Service System	III-9
3. Service System Notation	III-10
4. A Simplified Specific Service System for Initial Analysis and Example Computation	III-13
C. Flow Graphs for Measures of Effectiveness	III-25
1. Steady State Model	III-25
2. System Unsaturated Time	III-32
3. System Saturated Time	III-33
4. Waiting Time	III-34
5. Time in System	III-36

III	UNIT MODEL (Continued)	
D.	Computational Approaches and Sample Solutions	III-42
1.	Steady State (Equilibrium) Distributions	III-42
2.	Unsaturated, Saturated, and Waiting Time Distributions--Illustrative Results	III-56
E.	Areas for Further Research	III-60
1.	Numerical Calculations	III-60
2.	Extension of Model to Include Nonexponential Statistics	III-61
3.	Addition of Classification and Hand-off Function	III-62
IV	ASW FORCE EFFECTIVENESS MODEL	IV-1
A.	Model Development	IV-1
B.	Illustrative Results	IV-26
C.	Directions for Continued Methodological Development	IV-33
1.	Variable ASW Unit Motion	IV-34
2.	Intermittent and/or Combined Stimulus Presence	IV-34
3.	Intermittent Sensor Operation	IV-35
4.	Contact Hand-Off Process	IV-35
5.	Other Areas for Analysis	IV-36
V	REFERENCES	V-1

APPENDIX

A	FLOW GRAPHS AND SEMI-MARKOV PROCESSES ASSOCIATED WITH A MULTIPLE PHASE QUEUING SYSTEM	A-1
1.	Transition Rate Diagrams	A-3
2.	Markov and Semi-Markov Processes	A-4
3.	Competing Processes	A-4
4.	Multiple Phase Service Queuing in Terms of Competing Processes	A-5
5.	Flow Graphs for Markov and Semi-Markov Processes . .	A-7
6.	Using Semi-Markov Flow Graphs to Compute Probability Distributions for State Occupancy Times	A-17

APPENDIX

B SOLVING THE SYSTEM PROBABILITY EQUATIONS IN THE TRANSIENT STATE	B-1
1. The Differential Equations	B-3
2. The Successive Overrelaxation Iterative Method	B-7
3. Using the Flow Graph Structure to Organize Calculations Efficiently	B-11
DISTRIBUTION LIST	D-1

DD FORM 1473

ILLUSTRATIONS

I-1	Schematic Diagram of Study Approach	I-3
II-1	Typical ASW System Equipment/Functional Relationships	II-3
II-2	Examples of ASW Symbolic Configurations	II-8
II-3	ASW Force Functional Representation	II-10
II-4	ASW Force Functional Representation (Modified)	II-12
III-1	Prototype Queuing Representation of an ASW System	III-3
III-2	Alternative Representation for Figure III-1	III-4
III-3	Graphical Queuing System State Representations	III-12
III-4	Service System State for Idle Condition	III-16
III-5	Service System States with One Customer Present	III-16
III-6	Service System States with Two Customers Present	III-17
III-7	Service System States with Three Customers Present . .	III-19
III-8a	Service System States with Four Customers Present and No Internal Queues	III-21
III-8b	Service System States with Four Customers Present and Internal Queues	III-23
III-9	Steady-State Matrix Flow Graph for the Unit Model . . .	III-27
III-10	Truncated Matrix Flow Graph	III-28
III-11	"Folded" Infinite Chain in Steady-State Flow Graph . .	III-28
III-12	Calculation of Infinite "Tail" Value in a Flow Graph	III-30
III-13	Unsaturated Time Matrix Flow Graph	III-32
III-14	Saturated Time Matrix Flow Graph	III-34
III-15	Waiting Time Flow Graph for Random Selection from Stimulus Queue	III-36

III-16	Waiting Time Flow Graph for First-Come-First-Serve Policy at Stimulus Queue	III-37
III-17	Repetitive Waiting Time Flow Graph for First-Come-First-Serve Policy	III-38
III-18	The Four Copies of State 52 of Figure III-8b with Position of Specified Contact Designated	III-40
III-19	Detailed Diagram for 2 in System with Position of Specified Contact Designated	III-43
III-20	Partial Details of Diagram for 3 in System with Position of Specified Contact Designated	III-45
III-21	Partial Details of Diagram for 4 in System with Position of Specified Contact Designated	III-47
III-22	Time-In-System Flow Graph	III-49
III-23	Steady-State Probability Densities for Various Truncation Levels of Stimulus Queue	III-51
III-24	Steady-State Probability Densities Using Various Values of Input Rate	III-52
III-25	Configuration of Flow Graph for Up to Three in System Exhibiting Periodicity of Loop Transitions Having a Period of Five Steps	III-57
III-26	Probability Distributions for Unit Model	III-59
IV-1	Simplified Diagram of Sensor and Stimulus Relationships	IV-7
IV-2	Representative ASW Force for Model Demonstration . . .	IV-27
IV-3	Illustrative Force Effectiveness Model Results . . .	IV-31
A-1	Typical Flow Graph and Corresponding Equations . . .	A-8
A-2	Transmission Through a Flow Graph	A-9
A-3	Steady-State Flow Graph of a Markov Process	A-10
A-4	Matrix Laplace Transform Flow Graph for a Markov Process	A-12
A-5	Scalar Laplace Transform Flow Graph for a Markov Process	A-12
A-6	Matrix Laplace Transform Flow Graph for a Semi-Markov Process	A-15

A-7	Scalar Laplace Transform Flow Graph for a Semi-Markov Process	A-15
A-8	Semi-Markov Flow Graph for a Markov Process	A-17
A-9a	Semi-Markov Flow Graph for Time Spent in a Subset S of States	A-18
A-9b	Equivalent Flow Graph for Time Spent in in a Subset S of States	A-21
B-1	Block Diagonal Structure of the One Matrix A	B-6
B-2	Details of One Block Matrix in Figure B-1	B-7

TABLES

II-1	Definition of ASW Functions	II-2
II-2	ASW Functional Component Definitions	II-5
II-3	ASW Functional Symbology	II-7
III-1	Numbers of States in Subgraphs for Given Number of Contacts in System	III-41
III-2	List of States Occupied as Steady State Iterations Proceed Starting at State Number 1 (0 in System)	III-54
IV-1	ASW Unit Positions and Velocity Vectors	IV-26

GLOSSARY OF TERMS

ARRIVAL PROCESS -- A random process describing the probability distribution of interarrival periods for customers entering a queueing system.

COMPETING PROCESS MODEL -- A Semi-Markov process representation that phrases it in terms of independently operating stochastic processes which behave like random time clocks, the first of which goes off with an alarm being the one whose time is used to determine the next transition interval for the Semi-Markov process.

CONTACT -- A stimulus which has been detected by an ASW unit or force and is engaging the attention of an ASW crew.

CUSTOMER -- An object desiring service by the service system of a queueing system and having to possibly wait in the queue of the system.

DYNAMIC SYSTEM -- Term used in systems analysis to differentiate a time-varying model from a static one, i.e., a model with states changing in time and consideration of mechanics of their changes.

FLOW GRAPH -- A graphical representation of a set of linear equations, be they algebraic, differential, or whatever. The nodes of the graph correspond to the variables and the arcs correspond to the linear relations between the variables, the linear coefficients being the weight factors associated with the arcs of the graph.

MARKOV PROCESS -- A time series of change events with the property that the probabilities of what will happen next (i.e., what state transitions can occur and with what probability) are independent of whatever happened before except for whatever is happening now, i.e., only the present state determines the future.

QUEUE -- A mathematical model for a waiting line.

3

QUEUING SYSTEM -- A mathematical model for a queue along with a service system.

SATURATED SYSTEM TIME -- The duration when an ASW system is overburdened by processing contacts already obtained and is incapable of any further detections of stimuli.

SEMI-MARKOV PROCESS -- A time series of events very much like a Markov process except it is more general in the sense that holding times prior to new transitions are dependent on where the transition is going to be to, as well as what state the system is in now. In a Markov process, the holding times at the states are determined by knowing only what state the system is in now. For a precise definition, see Appendix A.

SERVER -- A facility for service in a service system of a queuing system.

SERVICE DISCIPLINE -- A set of rules for picking the next customer to be served out of a queue.

SERVICE PHASE -- One out of many (possibly parallel) types of service facilities that a customer passes through before it can be declared as "served" in a queuing system.

SERVICE STAGE -- A particular level of penetration of a customer into a service system. (In the ASW model, the first stage is called detection, the next stage is localization, etc.)

SERVICE STATION -- Another term for server.

SERVICE SYSTEM -- An arrangement of servers such that customers are required to pass through the servers in a specified series of service stages, and at each service stage the customer becomes engaged in another service phase.

STEADY STATE -- This refers to equilibrium (static) conditions as opposed to transient (dynamic) conditions.

STIMULUS -- An ocean phenomenon that could potentially result in the "detection" of a contact by an ASW unit or force. The phenomenon may occur due to the presence of a submarine or due to biological sources, acoustic propagation anomalies, or bottom and surface acoustic reflections.

TRANSIENT STATE -- This refers to a continually changing situation before settling down to an equilibrium.

TRANSMISSION
(OF A FLOW
GRAPH) -- The factor that one obtains as a multiplier for a flow graph. This is obtainable by algebraically eliminating all other variables in the linear equations represented by the flow graph.

UNSATURATED SYSTEM
TIME -- The duration of a period when an ASW system is not overburdened and is available for detection of a stimuli.

WAITING TIME -- The duration that a particular stimulus is theoretically present within detection range but is not detected because the ASW detection system is occupied.

I INTRODUCTION

The purpose of this research is to establish a methodology that can be used to evaluate ASW force effectiveness through the application of the queuing theory. The methodology represents the fundamental ASW functions (detection, classification, etc.) and the decision rules for false/true target evaluation, and force allocation within a time referenced framework, and specifies the effects of system contact prosecution and recovery times. The goals of the research are to:

- Establish measures of effectiveness of ASW systems that take into account their dynamic aspects
- Develop queuing models for individual ASW units and ASW force operations with the capability of accepting experimental and fleet-operational data
- Evaluate selected single ASW units and force models on the basis of fleet-operational data to demonstrate the utility of the methodology.

During the course of the study it was found to be expedient to extend the scope of the study beyond methodology conceptualization to include derivation of specific mathematical formulations and illustrative computations of various measures of system effectiveness. The scope of the project was extended in another manner. Originally, it was intended to consider the system capabilities of the individual ASW units in an aggregated sense. It quickly became apparent that this procedure would not suffice as an adequate representation. Accordingly, the internal functioning of an individual ASW unit was investigated in greater detail. The net result of this investigation was the development of the prototype unit model discussed in detail in Section III.

A schematic diagram of the study approach is presented in Figure I-1. Each of the elements of Figure I-1 is discussed in the report sections that follow with the exception of the first element: the real world of operational ASW. The definition of ASW functions, the establishment of the corresponding queuing analogy, and the construction of a force symbology and operational algebra are examined in Section II. Section III derives the appropriate unit measures of effectiveness and formulates the prototype ASW unit model. The force level measure of effectiveness and effectiveness model are described in Section IV. Supporting and background mathematical developments for the unit model are presented in the appendices.

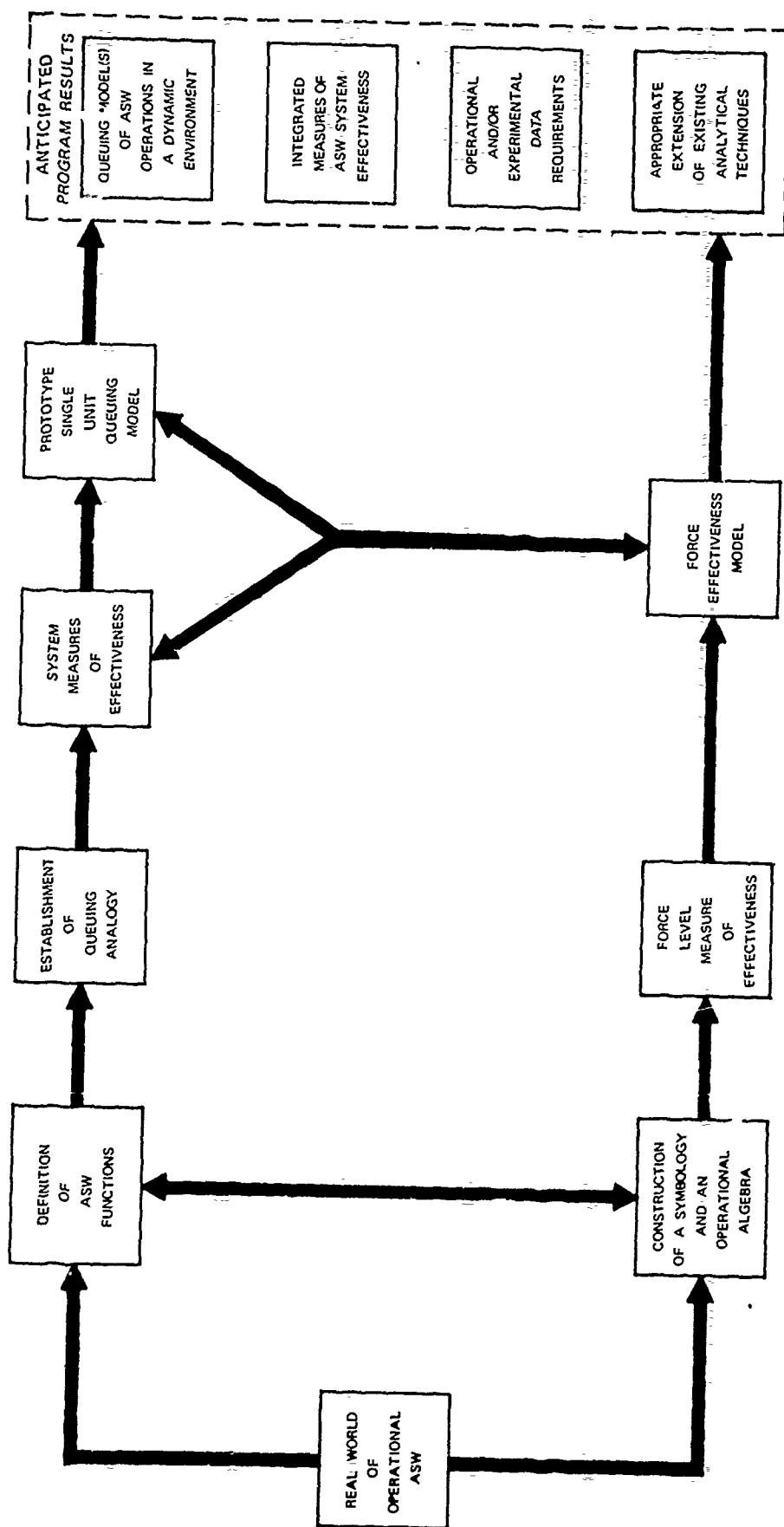


FIGURE I-1 SCHEMATIC DIAGRAM OF STUDY APPROACH

II GENERAL SITUATIONAL FORMULATION

A. ASW Functional Definitions

To be adequate, a system description should be applicable to a large number, if not all, of the different types of ASW systems, both in existence and postulated, and should be in terms that can be translated into queuing equivalents. There are many ways in which an ASW system may be described. The one used here is based on ASW functions. A functional system approach provides an adequate description since, by its very nature, it is sufficiently general to be applicable to all ASW systems; it is also appropriate in that the functions, if properly defined, can be represented as essentially discrete actions, which facilitates their translation into appropriate queuing equivalents.

On the individual ASW system level, five functional capabilities form the basic building block elements for formulating the queuing description of the system: detection, localization, attack, kill and classification. At the ASW force level, a sixth function is introduced, i.e., command and control.

The definitions, interpretations, and queuing equivalents of the six ASW functions are presented in Table II-1. Of the six functions, four are readily defined in terms of specific, essentially one-time actions and have been characterized as "servers" in queuing terminology. These four are: detection, localization, attack, and kill. The other two functions are continuous throughout the contact prosecution process and have been tentatively characterized as "system or force control functions" (queue discipline, contact processing rules, platform assignments, etc.)

Table II-1
DEFINITION OF ASW FUNCTIONS

ASW Function	Definition	Interpretation	Queuing Equivalent
Detection	Perception of contact stimuli in ocean environment	Enters or re-enters contacts into system	Server/Facility
Localization	Positioning of contact within operational coordinate system	Retains contact in system	Server/Facility
Attack	Determination of contact velocity vector sufficient for effective weapon launch and maneuver of weapon element into position adequate for weapon launch		Server/Facility
Kill	Actual weapon launch and destruction of contact	Eliminates contacts from system	Server/Facility
Classification	Decision as to validity of contact as a target	Controls contacts in system <u>Initial</u> --decision to admit contact to system <u>Subsequent</u> --decision to retain contact in system and/or to establish contact priority <u>Identification</u> --decision regarding identity of re-acquired contact	Establishes contact priority May reject contact at any stage prior to kill Influences system queue discipline
Command & Control	Allocation and control of ASW resources available to force	Controls contacts within overall Force System	Establishes time-dependent force organizational form. Defines operational queue system discipline. May eliminate contacts from system by avoidance

For the initial unit analysis, primary emphasis has been placed upon modeling the interactions among the four server functions under the assumption that internal system queue discipline is of the first-in-first-out type. This is equivalent to assuming that the system does not encounter false targets and all targets are equally threatening. The effect is to eliminate the classification function from the model formulation. This was done to constrain the complexity of the model until greater understanding is achieved about the interactions among the four server functions.

When the equipment/functional relationships in an ASW system are considered, the effects of the multipurpose character of many of the equipments quickly become apparent. For example, the nature of the actual equipment or resource allocations among the four server functions within a typical ASW system is shown in Figure II-1. In this figure, (X) signifies primary functional utilization and (/) implies secondary or continuing utilization.

Thus whenever a contact enters the system (upon completion of the detection function via the sonar equipment) at least a portion of the sonar capability must be dedicated to providing information about that

Function Equipment	Sonar	Combat Information Center	Fire Control	Platform Maneuver	Weapon Launch
Detection	X				
Localization	/	X			
Attack	/	/	X	X	
Kill/Re-attack	/	/	/	/	X

FIGURE II-1 TYPICAL ASW SYSTEM EQUIPMENT/FUNCTIONAL RELATIONSHIPS

contact in order that it may be properly processed. The detection capability of the system is correspondingly reduced. Hence, the extent of the system's detection capability is limited by the information rate of the sonar. As can be seen in Figure II-1, similar conditions apply to the other sensor functions. The implications of these effects will be discussed further in Section III where the formulation of the single ASW unit model is described.

B. Functional Symbolic Algebra

Having defined the ASW functions as the model building blocks for the single system, there exists a need to structure an aggregation of single systems. To satisfy this requirement, three ASW functional components are established and defined, and a symbology is constructed to describe the component interrelationships.

The three ASW functional components and their definitions are listed in Table II-2. In the functional hierarchy listed in Table II-2, the element and the operating element constitute the basic units from which specific ASW forces may be assembled. It should be noted that a platform need not possess all of the ASW functional capabilities to be classified as a unit. The difference between a unit and a force is defined in Table II-2.

Obviously, the nature of a force may change during the conduct of an operation due to a number of circumstances. Chief among these are: (1) reallocation of available resources, (2) attrition, and (3) temporary inhibitions of unit functional capabilities. In this request, the terms "uninhibited" and "inhibited" refer to primary and secondary utilization of a particular functional capability. For example, a destroyer may be operating with an ASW helicopter in such a manner that the attack and kill capabilities of the destroyer are subordinated to those of the

Table II-2
ASW FUNCTIONAL COMPONENT DEFINITIONS

ASW Functional Components	Definition
Unit	A single platform (ship, aircraft, helicopter, submarine) which possesses one or more complete or partial functional ASW capabilities
Operating Unit	A <u>unit</u> which is employed in a manner such that one or more of its functional capabilities are inhibited
Force	Two or more <u>units</u> and/or <u>operating units</u> which, collectively, comprise sufficient uninhibited functional capability for mission accomplishment--where mission accomplishment is defined as all necessary ASW functions from detection to kill

helicopter. In this case, the attack and kill capabilities of the destroyer do not cease to exist, but, insofar as the force is concerned, they are "inhibited" in favor of the helicopter capabilities. Thus the nature of the unit may change to that of an operating unit, and vice versa.

A force is not completely characterized by simply listing the assigned units and operating units. An essential ingredient is the operational doctrine and procedures that comprise the management or organization of the force. Clearly, the same units and operating units may be assembled and directed in many ways; each one of which constitutes a force with a differing level of effectiveness. To facilitate the description of these many possible force structures, an ASW functional symbology has been

developed as shown in Table II-3. The vocabulary is developed only to a stage which is consistent with the scope of the current effort and which demonstrates procedure and its application.

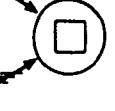
The majority of the symbols contained in Table II-3 have either been previously explained or are self-explanatory. Several examples of the use of these symbols are given in Figure II-2.

In Figure II-2A a detection operating unit is controlling an attacker unit. This example illustrates the use of the subscript to indicate a capability which is required of a unit or operating unit in order to adequately apply the primary or uninhibited capability. Only the four server capabilities are shown as inhibited or uninhibited. The classification function, since it affects all of the server capabilities, is indicated by the "C" subscripting the last uninhibited capability. To avoid unnecessary complexity, certain obvious required (subscripted) capabilities may be omitted from the unit and/or operating unit symbols. As an example, the required localization capability is omitted from the attacker element of Figure II-2A on the assumption that a knowledgeable observer will recognize the requirement without the necessity of including the extra symbol.

The examples of Figure II-2 are not just arbitrary collections of symbols. Each example is taken from a real world ASW situation. Thus Figure II-2A represents the functional organization of a destroyer (the operating unit) and a LAMPS helicopter team. Figure II-2B could be interpreted as an ASW helicopter unit (without weapon capability) working with a destroyer or ASW aircraft attacker operating unit. Figure II-2C represents the typical two destroyer search and attack unit (SAU). Figure II-2D depicts what may well be a future ASW configuration, i.e., two destroyers, or possibly two escort SSNs, each equipped with a form of the towed linear array system. In this case, localization will require the integrated performance of both operating units; hence the indication of localization

TABLE II-3

ASW FUNCTIONAL SYMBOLLOGY

SYMBOL	INTERPRETATION	SYMBOL	INTERPRETATION
D	A UNIT		INFORMATION AA TRANSFER (ONE WAY)
D, L, A, etc.	UNINHIBITED CAPABILITY OF A UNIT		INFORMATION TRANSFER (COORDINATION)
L ₀	SUBSCRIPT INDICATES A REQUIRED CAPABILITY IN CONJUNCTION WITH AN UNINHIBITED CAPABILITY		INFORMATION TRANSFER (DIRECTIVE)
	INHIBITED FUNCTIONAL CAPABILITY OF AN OPERATING UNIT (IN CONSONANCE WITH OPERATIONAL ORGANIZATION AND/OR PROCEDURES)		FUNCTIONAL INTERDEPENDENCE
D L A K	AN OPERATING UNIT		COMPETITION FOR SERVER

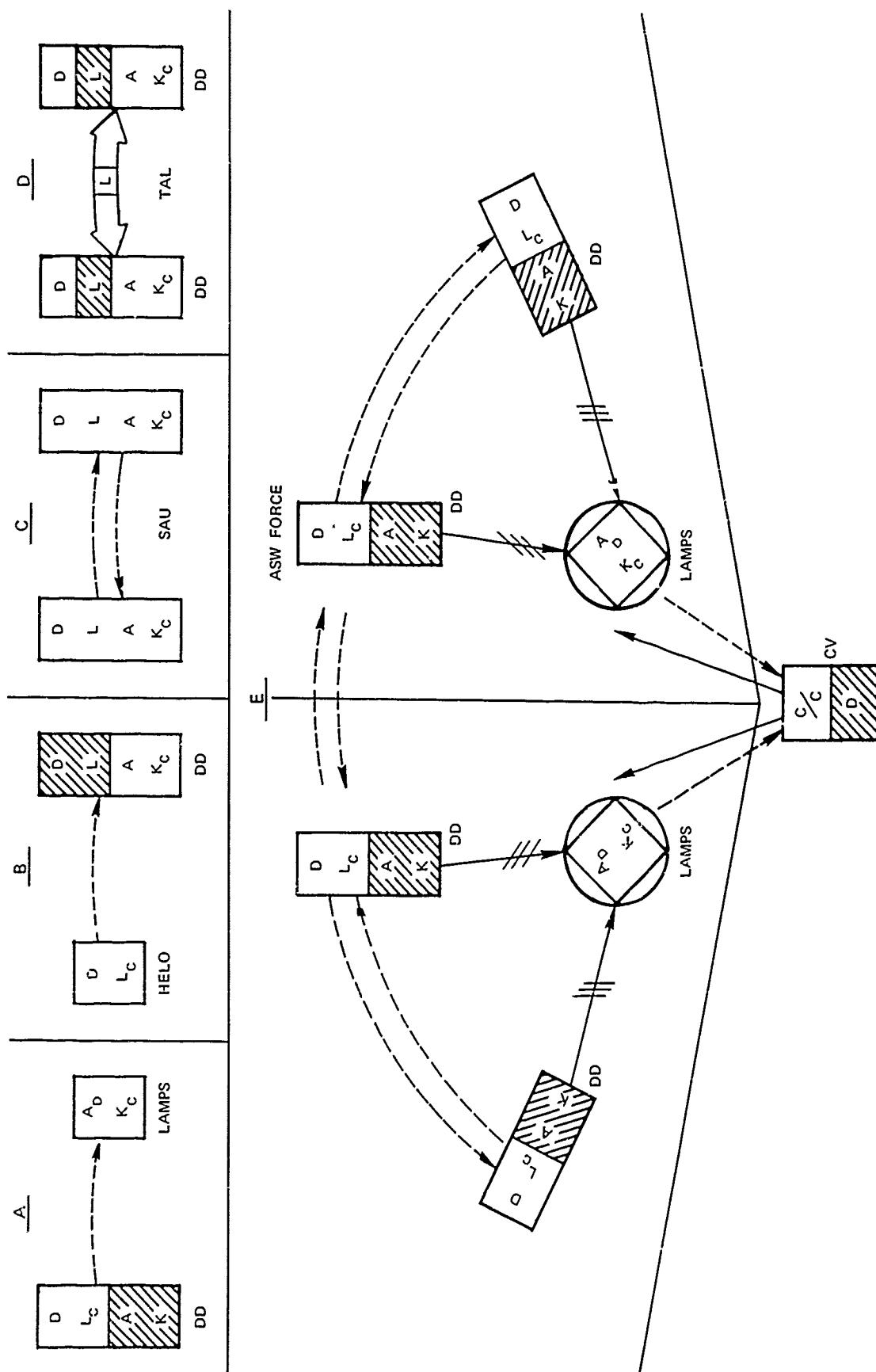


FIGURE II-2 EXAMPLES OF ASW SYMBOLIC CONFIGURATIONS

functional interdependence. It should be noted that all of the examples discussed so far constitute ASW "forces" in the context of the ASW functional components previously defined.

Figure II-2E illustrates the application of the symbology to a type of organization that has been associated with the term "ASW force." This force consists of four destroyer operating units organized into, primarily, two subgroups each with its own LAMPS attacker unit. In case of simultaneous destroyer contacts within a subgroup, there is a "competition" between the destroyers for the services of the attacker unit. The sixth ASW function, command and control (C/C in Figure II-2E), must now be introduced into the situation. The various command and control decisions that may be made in a situation such as this will not be dwelt upon here; they will be touched upon briefly in succeeding paragraphs. To complete the description of this specific example, the command and control operating unit of Figure II-2E can be interpreted as being an ASW aircraft carrier (CVS) or a carrier with an integrated air group (the CV concept).

The force representation of Figure II-2E can be misleading. To those familiar with ASW and, in particular, ASW formations, Figure II-2E may appear to represent spatial as well as functional relationships. While the various symbols may be positioned in some form of spatial order, this is not a requirement for the use of symbolic algebra. The purpose of the symbolic vocabulary and algebra is to display the ASW functional relationships within a given force. This point may be better appreciated by considering the ASW force shown in Figure II-3. In this figure the absence of any attempt to represent spatial relationships should be obvious, yet the functional relationships are clearly defined. By way of interpretation, the force shown in Figure II-3 represents a command and control operating unit, either a CVS or a CV, that is controlling four ASW aircraft (units A, B, C, and D), three ASW helicopters (units 1, 2, and 3), and two ASW groups

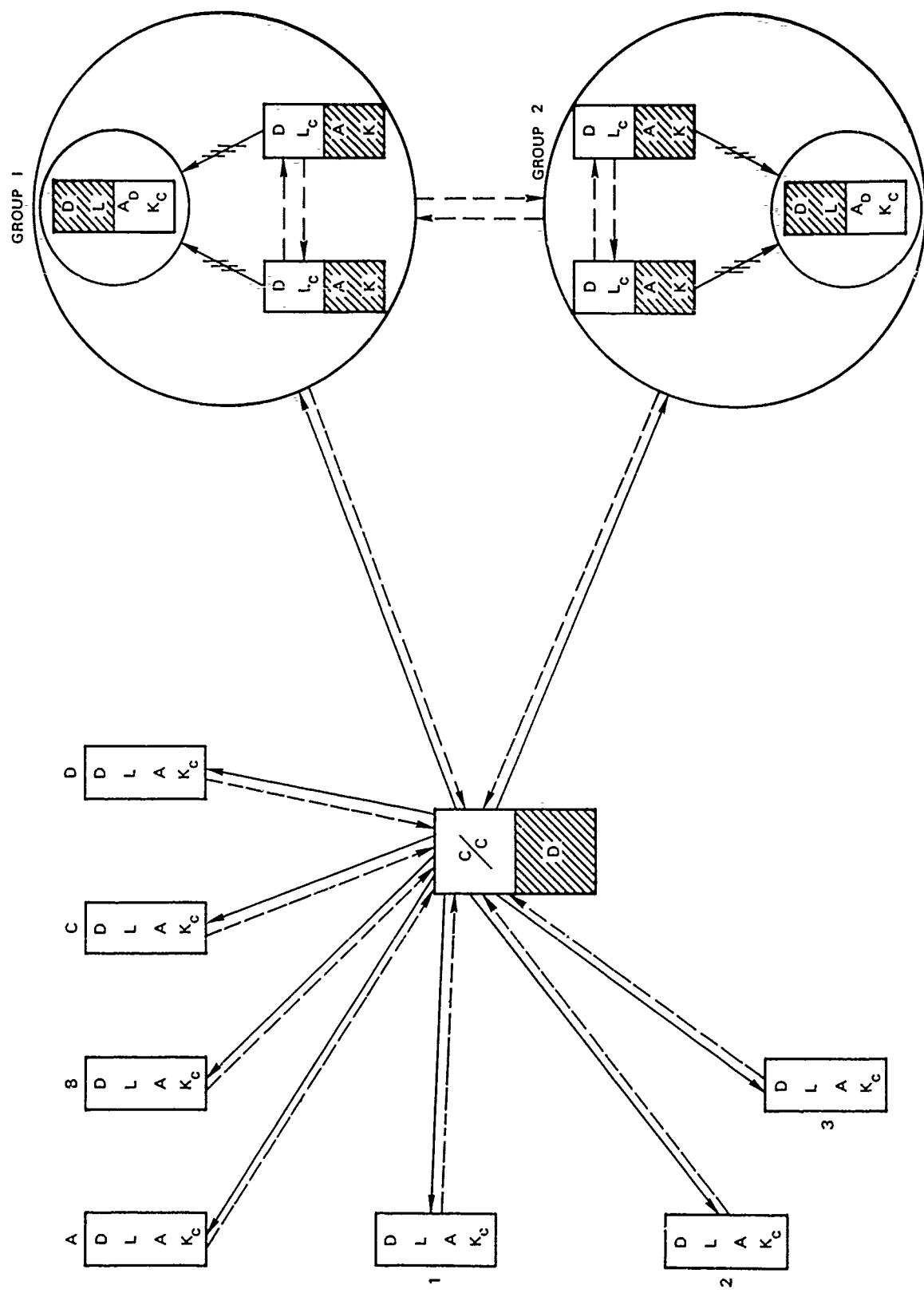


FIGURE II-3 ASW FORCE FUNCTIONAL REPRESENTATION

consisting of two destroyer detection operating units and an ASW helicopter attack operating unit. To avoid unnecessary complexity in Figure II-3, only the principal lines of information transfer are shown. The situation in Figure II-3 may also be considered to represent the force status during the period when only ASW search is being conducted. The actual force configuration and organization may take many forms, depending upon the tactical situation and/or the various command and control decisions implemented once a target is encountered. For example, the force organization may undergo the changes indicated in Figure II-4 after a target is detected. In this instance, the aircraft unit D has made contact. A command and control decision has been made to detach the attacker operating unit from group 1 and assign it to assist D. The remaining destroyers of group 1 have been retained, organizationally, as a group or SAU. Note that, in this case, the nature of each destroyer in group 1 changes from that of an operating unit to that of a unit. This brief example illustrates how the symbolic vocabulary and the associated operational algebra can be used to depict the time dependent functional capability changes an ASW force may undergo during a dynamic ASW engagement.

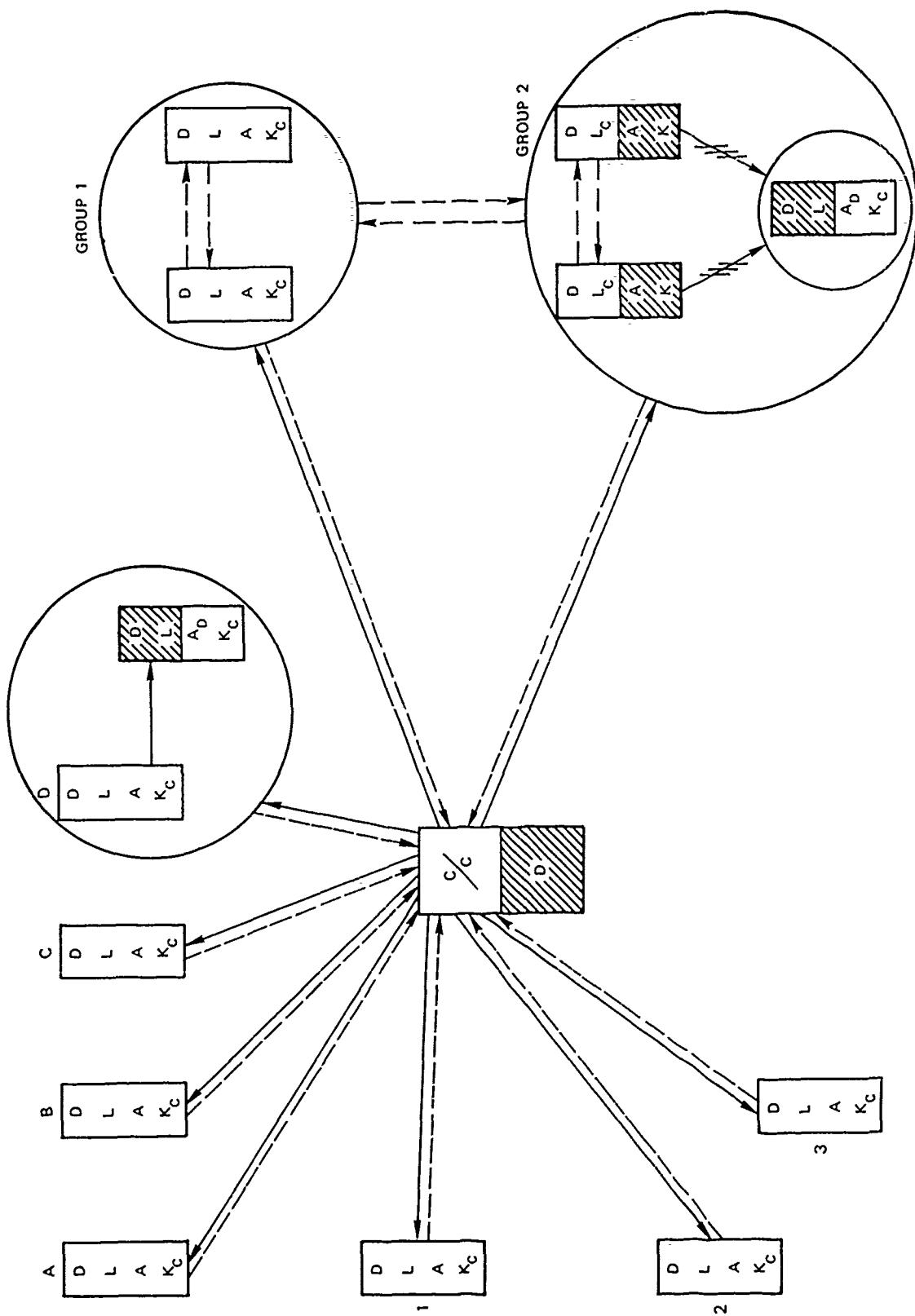


FIGURE II-4 ASW FORCE FUNCTIONAL REPRESENTATION (MODIFIED)

III UNIT MODEL

A. Introduction

1. Description

For the unit ASW model it is assumed that, as a contact passes through the system, it occupies a unit of sonar service while in the first "server," i.e., while undergoing detection. Throughout its stay in the system it continues to occupy that unit of sonar service, even when it is in any of the three queues following the first server. Similarly, the CIC resource is assumed to be occupied at the rate of one unit per contact from the moment of entering "localization" and up to leaving the system, etc.

Observe that we have been led to model discretely the capabilities of each functional server to service multiple contacts simultaneously. It is conceptually clear that there are limits on simultaneous functional service. Thus it is reasonable to establish capacities on simultaneous service. Nonetheless, establishment of definitive values for such capacities must await future attention. The representation used weighs all contacts uniformly as regards the load each places on a server, or equivalently, as regards the difficulty of service.

The command and control and classification functions have been omitted from the present model; they will be represented as the processes that determine priority of service and that specify alternative paths through the system networks (Section III-B). Similarly, certain other features of an ASW engagement have not yet been modeled. Some of these are: (1) hand-off of functions from other ASW units, (2) lost contact processes, and (3) unsuccessful kill attempts and reattack process.

2. Outline of Queuing Model

Based on the above considerations, the initial multiple queuing situation of Figure III-1 was selected for study. The meaning of Figure III-1 follows.

Acoustic stimuli, be they locally generated as they become available for detection, await the availability of the detection server D in a fictitious queue Q_D . This queue is created to measure the statistics of missed and delayed detection due to preoccupation of the detection process. As soon as an element of the detection server is available (for specificity we use here four degrees of detection server occupancy) the stimulus progresses "through" the detection phase according to prescribed detection time statistics. From this point the stimulus is a contact. Contacts not in process of localization are represented as "waiting" in a queue Q_L in the same manner as the contacts progressed from a queue awaiting detection to a queue awaiting localization. So, the contacts progress from the latter queue, through the localization process, through a queue awaiting the attack function, etc., through "service" by the kill function.

The number of parallel "service units" at each stage was selected for development purposes to be just enough so that queuing is possible at the following stage. That is, in view of service at each stage being "tied up" until a contact leaves the system, it is necessary to have at least one less unit of service at each stage as one progresses. In other words, we selected the server capacities to be simple enough for a first study, but complicated enough to allow queues to form in all possible stages.

An alternative schematic representation of the situation is presented in Figure III-2. This figure shows the idea of the presence of a contact being felt in several places at once, in that a contact

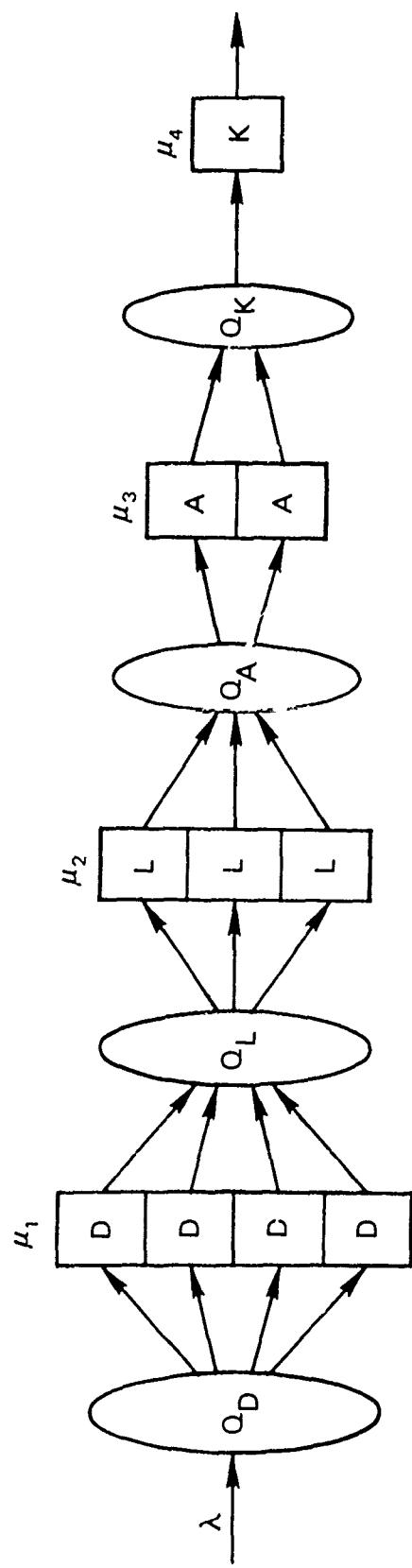


FIGURE III-1 PROTOTYPE QUEUING REPRESENTATION OF AN ASW SYSTEM

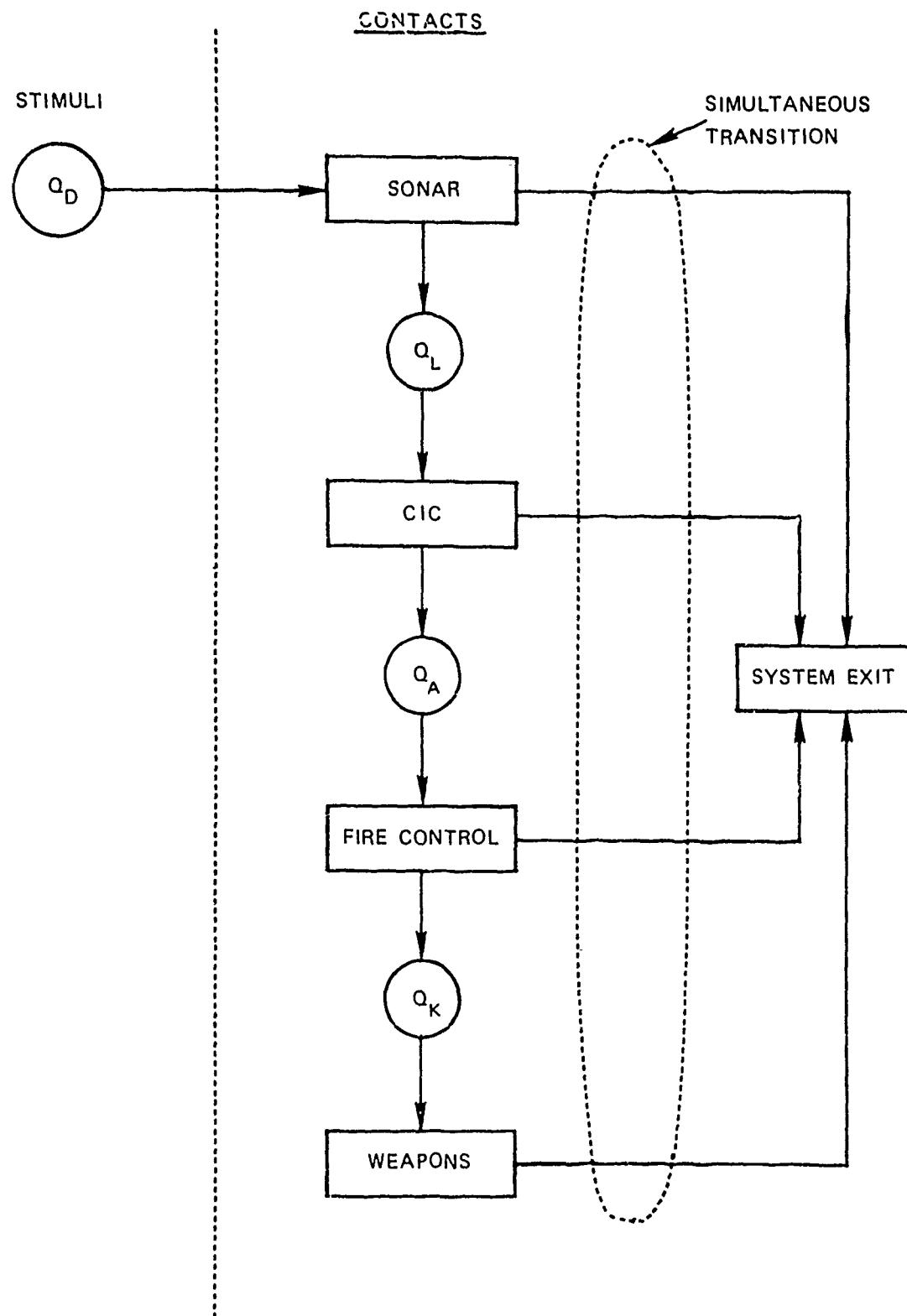


FIGURE III-2 ALTERNATIVE REPRESENTATION FOR FIGURE III-1

progressively uses the resource of sonar (detection), then sonar and CIC (localization), etc., until it finally exits from the system.

As a start it was assumed that all parallel service units at each "stage" have the same exponential "service" distribution, with parameters as in Figure III-1. It must be remembered that a "completion" of a service does not render a given service unit available for other contacts; such a completion merely tells when a transition of a contact to the next state occurs (either into the next "server" or the queue ahead of it). The input distribution for all true and false contacts is also assumed to be exponential with parameter λ . Both the input and service distributions can be made more general, using the so-called "method of stages," which amounts to setting up a network of parallel and series exponential holding times.

3. Measures of Effectiveness

Various attributes of factors associated with queuing systems in general were considered in structuring a suitable model for evaluating ASW system performance. The primary statistical factors were judged to be:

- Unsaturated time distribution (for entire system)
- Saturated time distribution (for entire system)
- Waiting time distribution (in first queue, i.e., ahead of detection)
- Distribution of time spent in entire system.

Other factors of significance are:

- Distribution of queue lengths at the various stages
- Distribution of the number of contacts exiting any stage in a given time.

The primary statistical factors are needed in the force effectiveness model described in Section IV. The others are useful for evaluating the effects of physical storage capacity or other limitations at various stages of the ASW system, and for determining the nature of the input distributions to other ASW platform queuing models coupled to a given one through a command and control hand-off system.

The above statistical factors lead to calculation of the following possible measures of effectiveness of an ASW unit:

- Probability that a contact is lost
- Expected number of contacts lost while in system, over a given period of time
- Probability of missing a true contact (failure to detect) for a given level of contact activity (false and true)
- Expected time to detect
- Expected time from detection to kill.

When the classification function is added to the model, the following additional measures will be obtainable:

- Probability of losing a contact before classifying
- Expected time to classify
- Probability that a false contact is pursued all the way through to the kill function.

Work to date was aimed only at numerical calculation of the statistical distributions as these are the most difficult to obtain. The above measures-of-effectiveness, which are relatively easy to obtain once the distributions are known, were not considered in the present effort so far, lacking fleet-operational data to render them meaningful.

4. Approach Selected for Model Analysis

Fortunately, a suitable approach to the ASW problem has arisen, in spite of the difficulties with attempts at simpler models. The model is involved but, on the other hand, it is computationally tractable and it does in fact yield all of the above measures of effectiveness as possible outputs. None of these measures is amenable to "closed form" analytical evaluation. Instead, numerical computation of formulae is employed. The model does provide illustrative procedures for displaying how these measures are evaluated. The developments in this approach were motivated by the doctoral work of Dr. Thomas E. Humphrey of Stanford Research Institute, on multidimensional Markov processes applied to queuing theory.⁵

A Markov model for the single unit situation is obtainable from the following procedure:

- (1) Assign a separate state of the system for every probabilistic event that it is desired to distinguish for the desired computations.
- (2) Establish the transition probabilities from one state to another and the distributions of duration in the various states.
- (3) Establish the appropriate semi-Markov flow diagram⁶ created by the resulting embedded continuous-time Markov process in the model.
- (4) Calculate the "transmission" (a Laplace transform) through the flow diagram between the input and output nodes associated with the desired measure of effectiveness. This is likely to be a large, but not infeasible computational task for the size of networks in the ASW case. Fortunately, the calculations are suitable for digital computation.
- (5) Calculate the inverse Laplace transform of the "transmission" in 4 above. This is equal to the cumulative distribution function of the quantities specified in the measure of effectiveness.

For present computations, a numerical integration approach was selected based upon ease of implementation. The procedure in steps 4 and 5 was not used since the Laplace transforms resulting from the very large flow graphs involved here are too complex to treat without a computer program for automatic flow graph reduction. Granted, it would be very illustrative to know the various time constants of the model (i.e., eigenvalues of the Markov transition matrix), but numerical plots of the probability distributions corresponding to the various desired measures of effectiveness are a suitable form of the desired information for most purposes. The Laplace transform approach has not been totally discarded, and a search is continuing for a practical method of solution by this means.

B. Detailed Description of Unit Model

1. Structure of the Queuing Situation in the Unit Model

The queuing situation to be studied is one in which each customer* is subjected to multiple phases of service by parallel servers according to a particular discipline. The customer may be required to wait between service phases and/or may leave before completing all service phases.[†] The methods developed are not limited to the case studied but are applicable to a broad class of queues having multiple phases of service (single or multiple servers) and various service disciplines, queue disciplines, and arrival behaviors. Poisson distributed arrivals and exponentially distributed service phase holding times are assumed, but the approach described is not limited to these distributions alone.

* Section IV-B will speak of a contact as a "customer." This is a very familiar term in general queuing theory and is more illustrative of the "queuing" concept than "contact."

† This corresponds to classification as a non-sub in ASW situation.

As pointed out at the end of Section III-A-2, it is possible to replace each exponential holding time by a subnetwork of holding times approximating any desired distribution. The queuing situation is considered to consist of an arrival process, an arrival queue, and a service system.

2. The Service System

The queue service system is one in which each customer is subjected to successive phases of service at a sequence of service stations according to a specific service discipline. The service stations are arranged into a sequence of service stages, as shown in Figure III-1. Each stage provides one phase of service. The customer moves from stage to stage in sequence, without repetition, and may leave the service system after completing any service phase.

A customer is served by one and only one service station at each stage. Parallel stations allow services up to s_k customers at stage k . The capacity of successive stages decreases, i.e., $s_{k+1} \leq s_k$. A customer completing stage k service may find no service stations available at stage $(k + 1)$ and be forced to join a stage $(k + 1)$ queue. The maximum length of the stage $(k + 1)$ queue is $s_k - s_{k+1}$. The stage 1 queue accommodates customers arriving at the service system and need not have a maximum length.

A station, having served a customer, remains dedicated to that customer until he leaves the service system. A customer receiving service at stage k or in the stage $(k + 1)$ queue "ties up" one service station in each of the first k stages; the available service capacity of each of the first k stages is reduced by one. When a customer leaves the service system, the stations that were dedicated to that customer are released and become available to serve new customers.

The maximum number of customers that can be in the service system at one time is the server capacity of the first stage, s_1 .

3. Service System Notation

Let n_k be the number of service stations in stage k that are busy, i.e., that are either serving customers or dedicated to customers elsewhere in the service system. Let q_k be the number of customers in the stage k queue. From the description of service system behavior, it follows that:

$$n_k \leq s_k \quad (\text{III-B-1})$$

$$q_k = 0 \quad \text{if } n_k < s_k \quad (\text{III-B-2})$$

$$0 \leq q_k \leq n_{k-1} - s_k \quad \text{if } n_k = s_k$$

$$n_k + q_k \leq s_{k-1} \quad \text{for } k > 1 \quad (\text{III-B-3})$$

The value q_1 describes the queue of arrivals to the service system and need not have a limit. The value n_1 describes both the number of busy stage 1 service stations and the total number of customers being served. The remaining stages of a k -stage service system may be characterized by the ordered N-tuple

$$\Phi = [n_2, \dots, n_k, q_2, q_3, \dots, q_k] , \quad (\text{III-B-4})$$

which takes a finite number of unique values. This allows the values of Φ to be ordered and identified by integer valued state numbers. The state

number and n_1 together describe the state of the full service system. The value q_1 gives the length of the arrival queue. The three numbers together characterize the state of the total queuing system.*

A more compact representation is possible. The total number of customers utilizing stage k service stations or in the stage k queue is given by the sum:

$$c_k = n_k + q_k . \quad (\text{III-B-5})$$

The value of c_k can be decomposed using the equations:

$$n_k = \begin{cases} c_k & \text{if } c_k \leq s_k \\ s_k & \text{otherwise} \end{cases} \quad (\text{III-B-6})$$

$$q_k = \begin{cases} 0 & \text{if } c_k \leq s_k \\ c_k - s_k & \text{otherwise} \end{cases} . \quad (\text{III-B-7})$$

The total queuing system can be fully characterized using the value c_1 and an ordered N-tuple consisting of the values c_2, c_3, \dots, c_k . Such

* This statement implies certain assumptions regarding the service stations in each stage. The probability of completing a customer service is assumed to be a function only of the number of stations currently serving customers and is independent of which stations are supplying the service. When the probability of completing a service is dependent upon which stations are supplying the service, a description of the service system state must specifically identify each busy service station with a particular customer. The service discipline must specify the rules by which idle service stations are assigned to customers. This is certainly possible using the methods developed here, but the number of unique service system stages is greatly increased.

N -tuples can be arranged in a one-to-one correspondence with the Φ ordering and assigned the same state numbers. The state of the total queuing system can thus be characterized by c_1 or the pair of values (N_1, q_1) , together with either the N -tuple Φ , the N -tuple of c_k 's, or the service system state number.

For mathematical formulations of queuing system models, and where concise notation is required, the value c_1 , together with the state number is most convenient, and will be represented by the notation (c_1, x) .

For graphical representations of service system state transitions, the n_1 and q_1 values and the N -tuple Φ will be used since this makes the system state highly visible. The values will be portrayed as shown in Figures III-3a and III-3b. The Figure III-3a representation is used only when all queues in the system are empty. The concise notation (c_1, x)

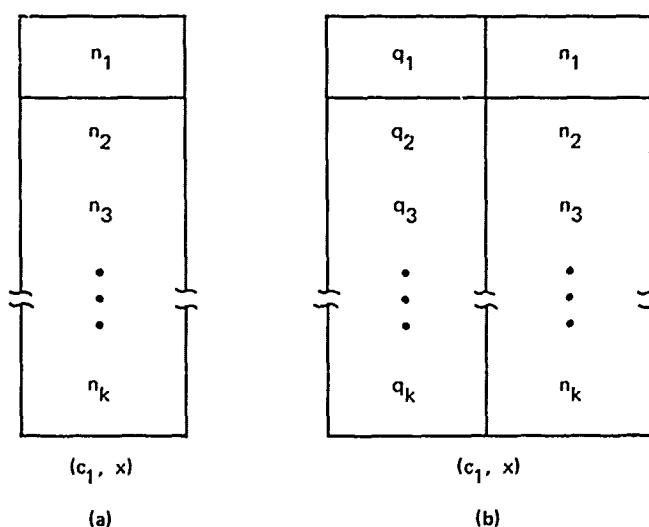


FIGURE III-3 GRAPHICAL QUEUING SYSTEM STATE REPRESENTATIONS

will be shown, where possible, immediately below the graphical representation to identify the corresponding state number assignment.

4. A Simplified Specific Service System for Initial Analysis and Example Computation

Discussions in this and subsequent sections are in terms of the particular service system configuration shown in Figure III-1. The ten service stations are arranged to provide four phases of service. A maximum of four customers can be accommodated by the service system at one time. The queue preceding stage 1--the arrival queue--is of indefinite length. There are four service stations for stage 1, three for stage 2, two for stage 3, and one for stage 4. The queues preceding stages 2, 3, and 4 each have maximum length one. This service system is readily generalized to service systems having different numbers of service stages and/or service stations at each stage.

It is assumed that all service stations in a stage are indistinguishable. A service station in stage k is assumed to have an exponentially distributed holding time with mean value μ_k . Thus if m_k is the number of stations serving customers in stage k , the probability that one of the stations will complete service in an interval of length dt is equal to $m_k \mu_k dt$. Note that $m_k = n_k$ only for the last stage in the service system, and is given by

$$m_k = n_k - n_{k+1} - q_{k+1} \quad (\text{III-B-8})$$

for all other stages.*

* It is possible to make service completion probability a function of any or all of the values m_k , n_k , and q_k to model situations in which server activity varies with customer load.

In the interval dt , one and only one service station may complete service and produce a state transition. That is, station service completions are considered to be independent, mutually exclusive events. When a service completion results in a customer's leaving the service system,* however, all service stations previously dedicated to that customer are freed to accept new customers. When such a transition occurs, a customer in a queue is allowed to begin service immediately. Thus any single service completion in which a customer leaves the service system causes transition to a state in which all interim queues--i.e., those queues associated with stages 2, 3, and 4--are emptied and the arrival queue, if nonzero, is reduced by one. New customers enter service stations freed by a departing customer simultaneously with the customer departure from the service system. While this represents a compound reallocation of service station resources, it results from a single station service completion. Compound state transitions corresponding to completion of more than one station's service in an interval of length dt do not occur.

State numbering is such that in transitions of the form $(c_1, X) \rightarrow (c_1, Y)$ it is always the case that $Y > X$. Further, the numbers (X and Y) are assigned so that the four states occurring with one customer in the system are numbered 1 to 4, the ten states occurring with two customers in the system are numbered 1 to 10, the 19 states occurring with three customers in the system are numbered 1 to 19, and the 28 states occurring with four or more customers are numbered 1 to 28. These conventions simplify the task of relating the graphical representations to mathematical formulations of the model. In other respects the state

* The simplified system here allows customers to leave the entire system only after the 4-th stage of service. Modeling of classification and hand-offs is nevertheless possible by reverting to a more general situation, as described prior to this subsection.

number assignments are arbitrary. Figures III-4 through III-8 identify and number the states and graphically describe state transitions for the service system configuration of Figure III-3. Three cases are identified: the idle server system, the busy service system, and the saturated service system.

The service system is idle when no customers are present. The single system state is represented in Figure III-4. No service system state transitions occur when the service system is idle.

A busy service system results when one, two, or three customers are present. The service system states are identified and numbered, and state transitions are represented graphically for these three conditions in Figures III-5, III-6, and III-7, respectively. For each state the c_1 value and state number assignment are shown in the notation (c_1, X) immediately below the state representation; service system states resulting from customers leaving the service system are also shown in this notation. Transitions representing customers leaving the system are shown only for stage 4 service completions. Associated with each arrow describing a state transition is a coefficient. This coefficient, when multiplied by dt , gives the probability that the transition will be made in an interval of length dt . Note that contention for service station resources results as soon as two or more customers are in the service system, producing system states having interim queues (Figures III-5 and III-6). For definitional purposes, an "unsaturated" system is one which is either idle (empty) or busy.

A saturated service system results when four or more customers are present. Because of the large number of possible service system states (28) in this case, the graphical representation of service system state transitions is in two parts: Figure III-8a describes service system states that exist when no interim queues are present; Figure III-8b describes

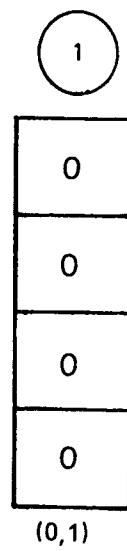


FIGURE III-4 SERVICE SYSTEM STATE
FOR IDLE CONDITION

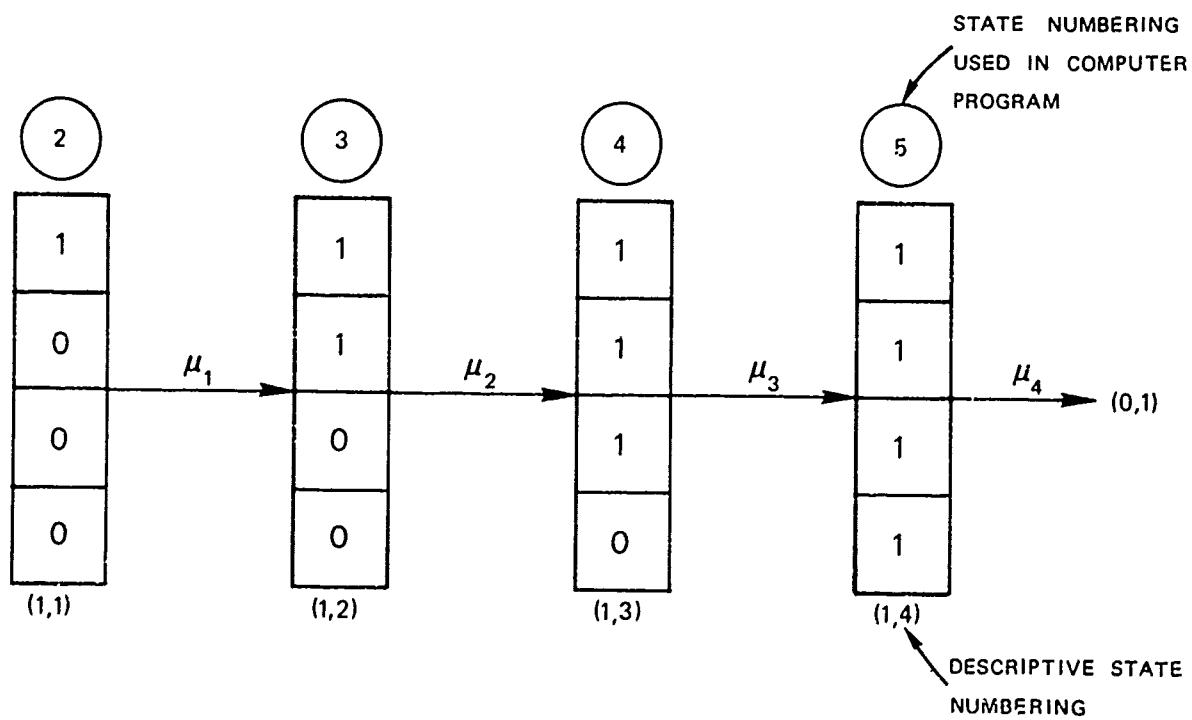


FIGURE III-5 SERVICE SYSTEM STATES WITH ONE CUSTOMER PRESENT

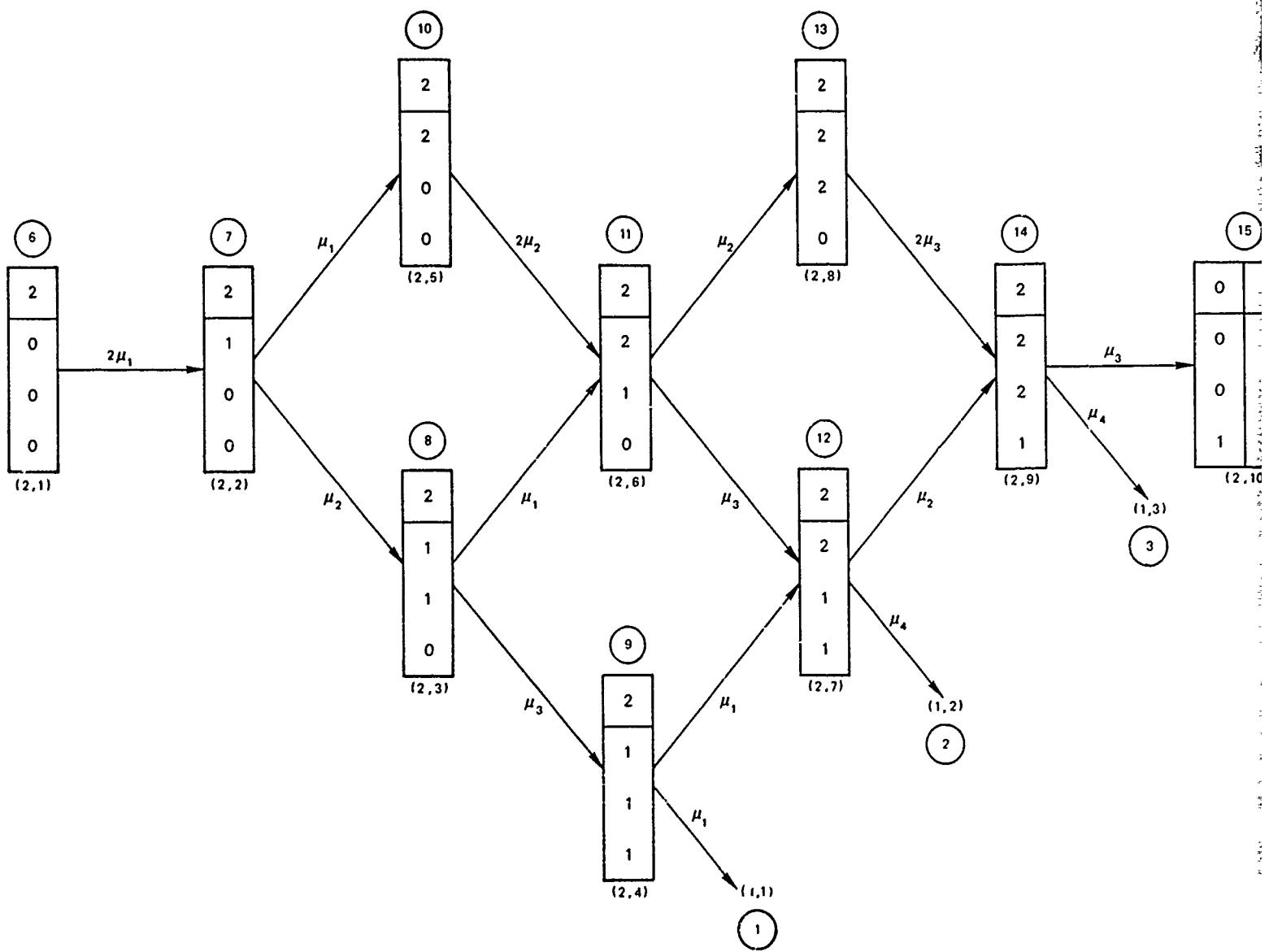


FIGURE III-6
SERVICE SYSTEM WITH TWO SERVERS PRESENT

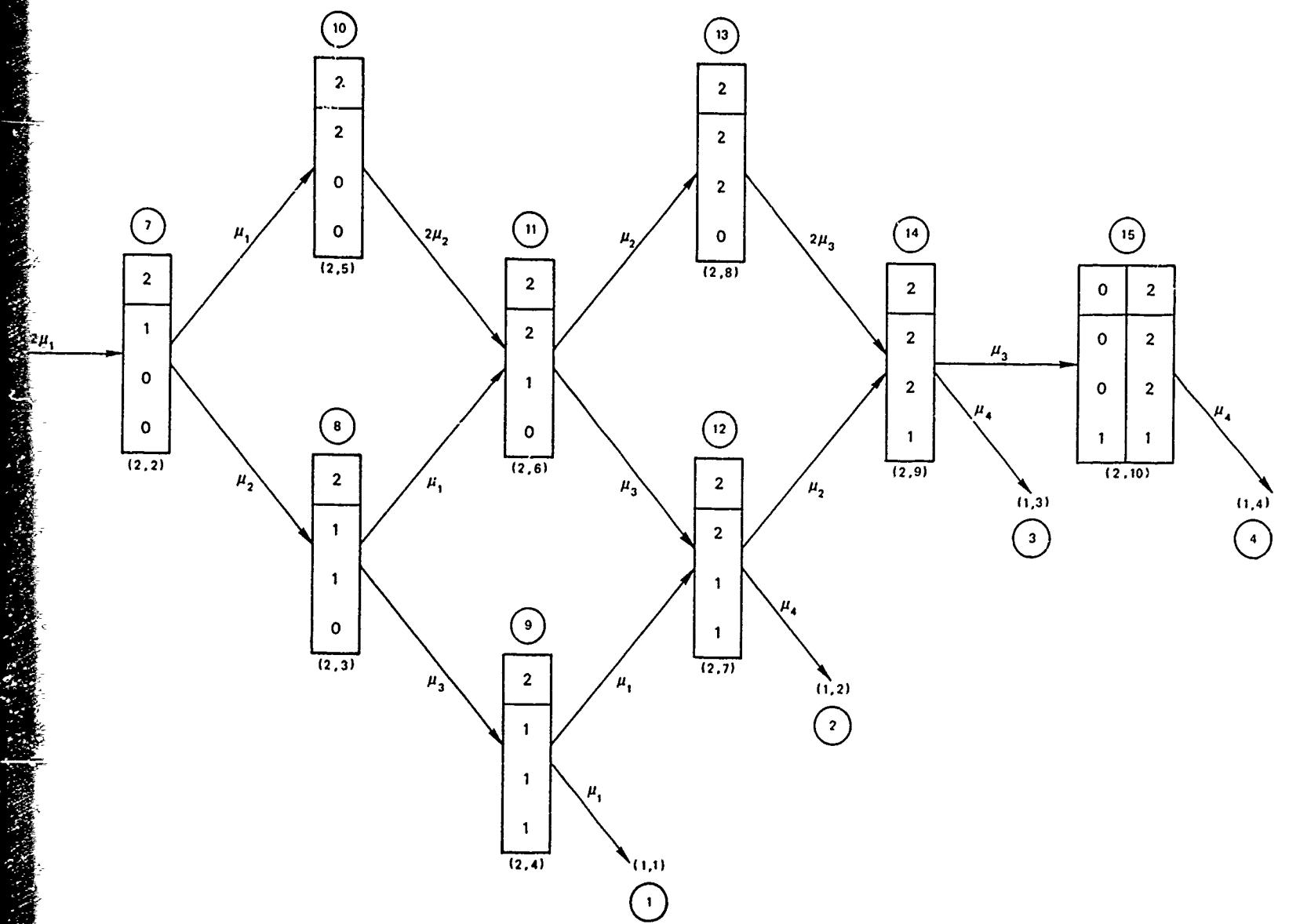
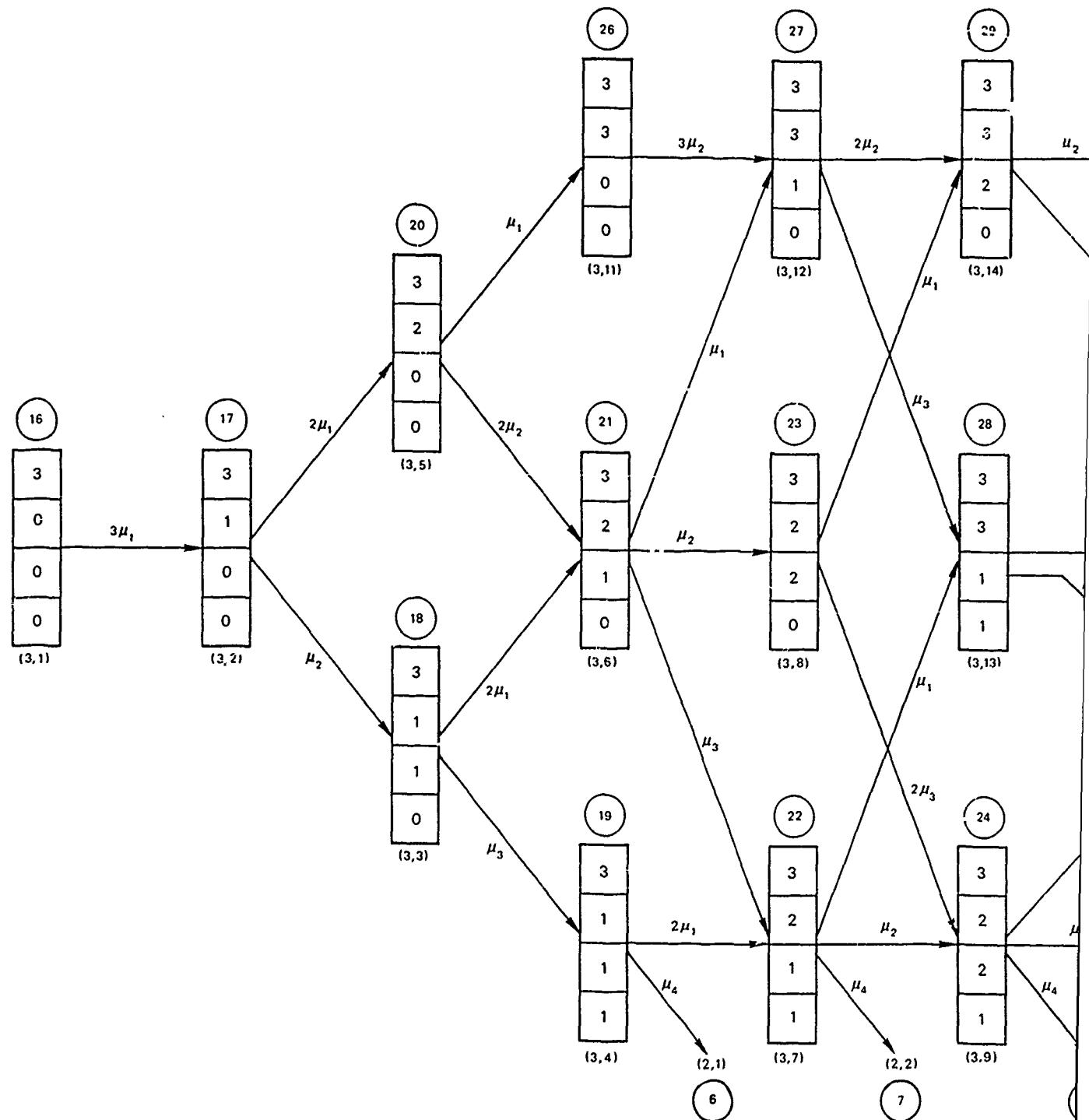
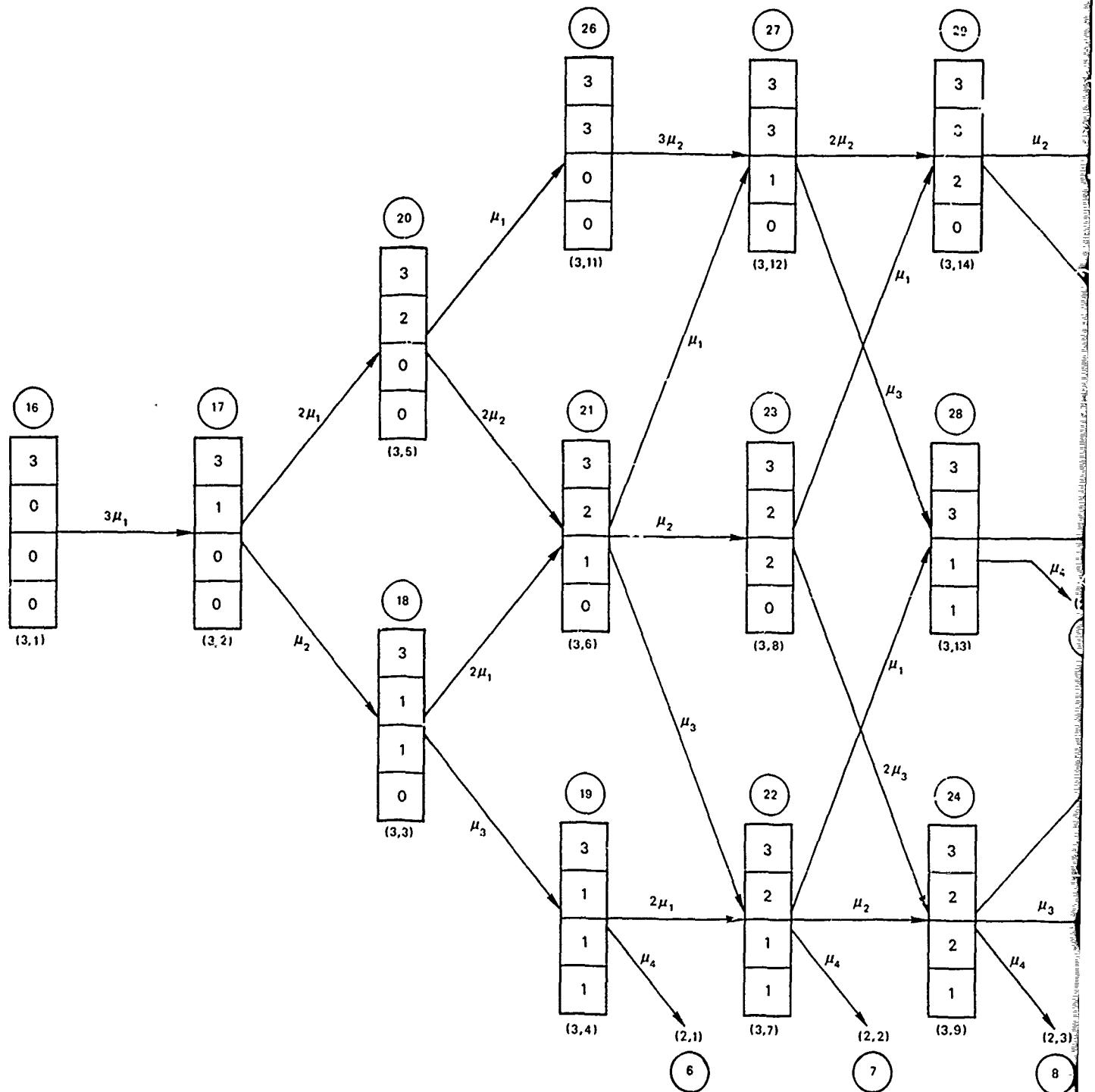


FIGURE III-6 SERVICE SYSTEM STATES WITH TWO CUSTOMERS PRESENT





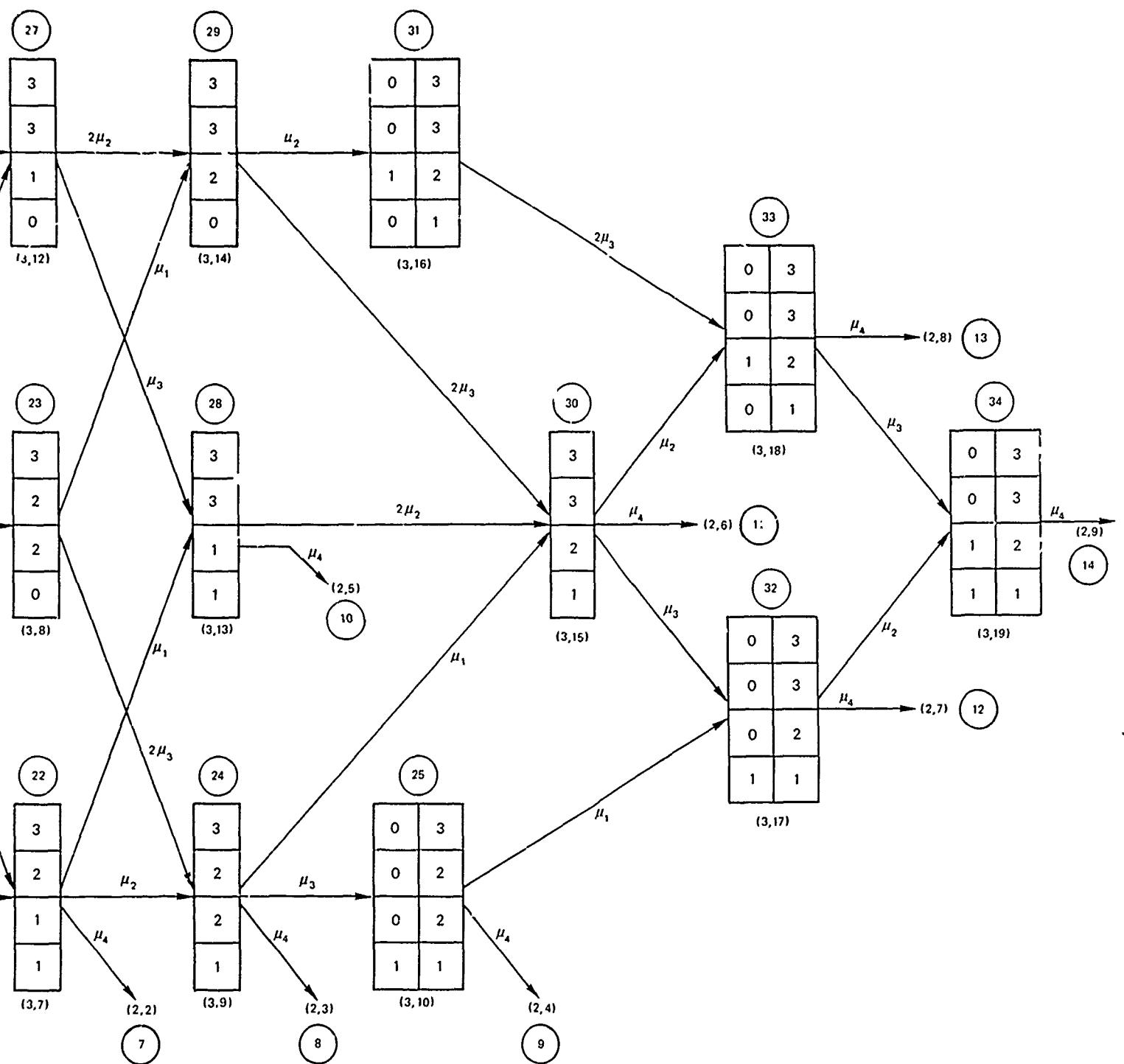


FIGURE III-7 SERVICE SYSTEM STATES WITH THREE CUSTOMERS PRESENT

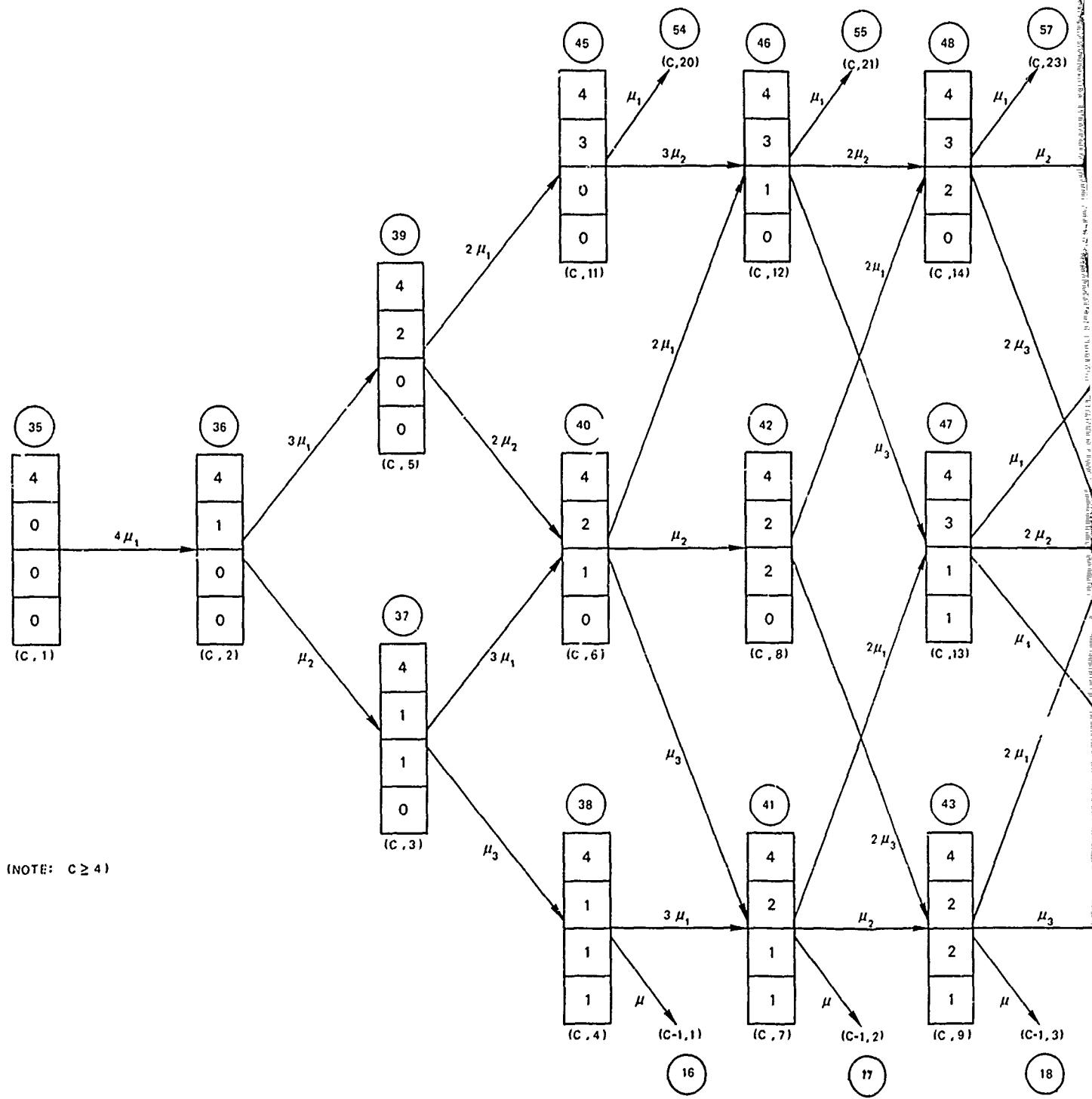


FIGURE III

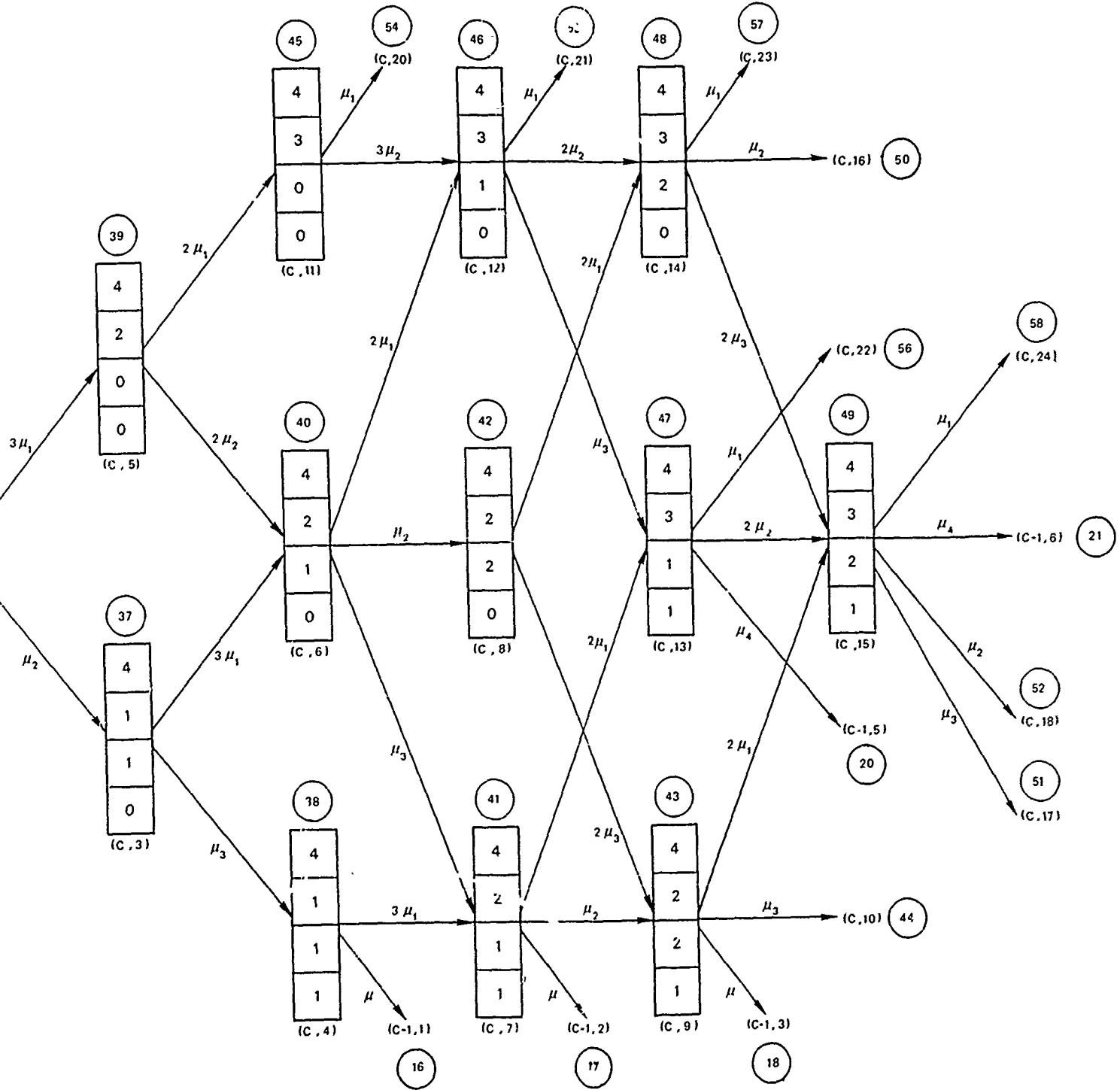
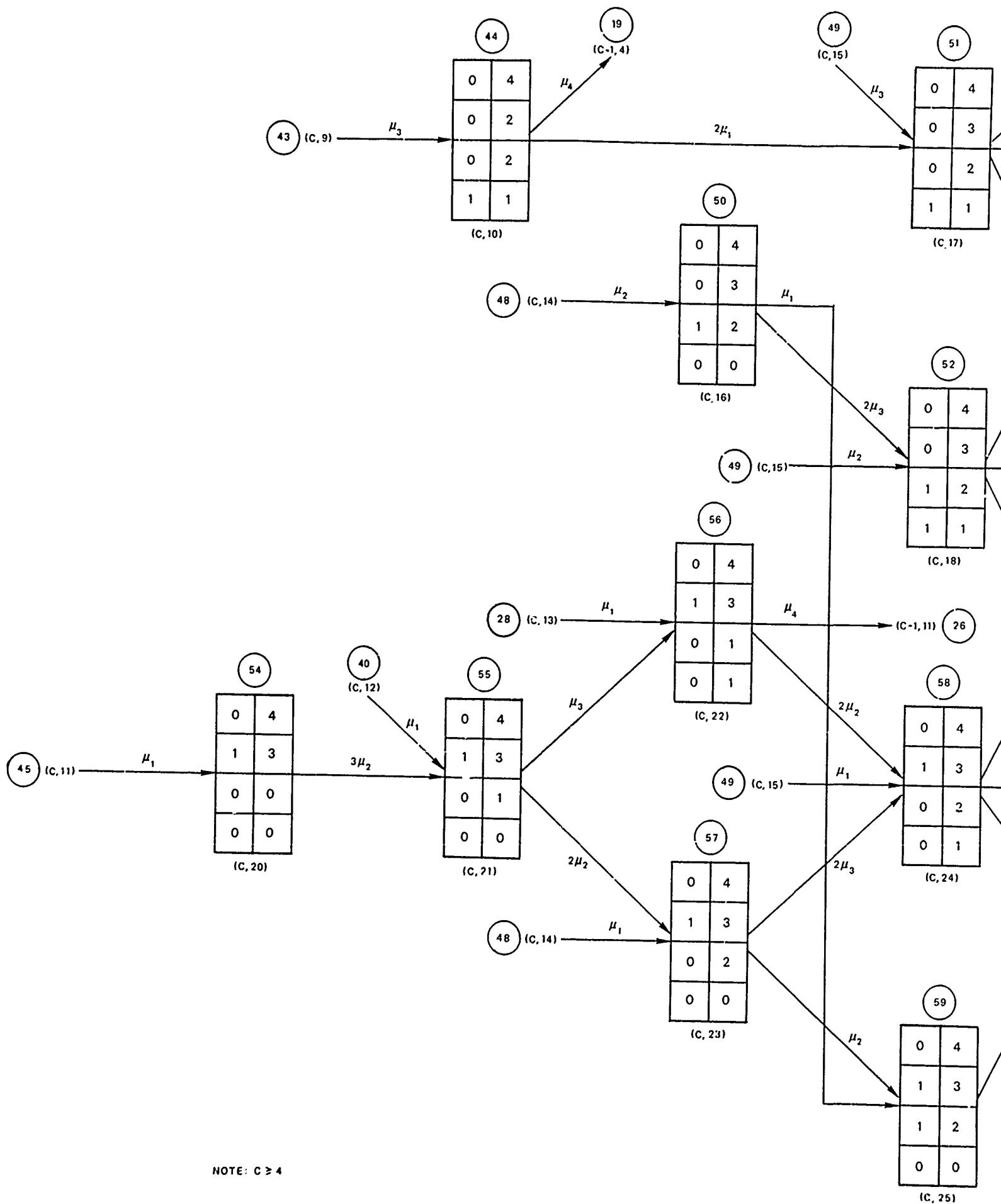


FIGURE III-8a SERVICE SYSTEM STATES
WITH FOUR CUSTOMERS
PRESENT AND NO
INTERNAL QUEUES



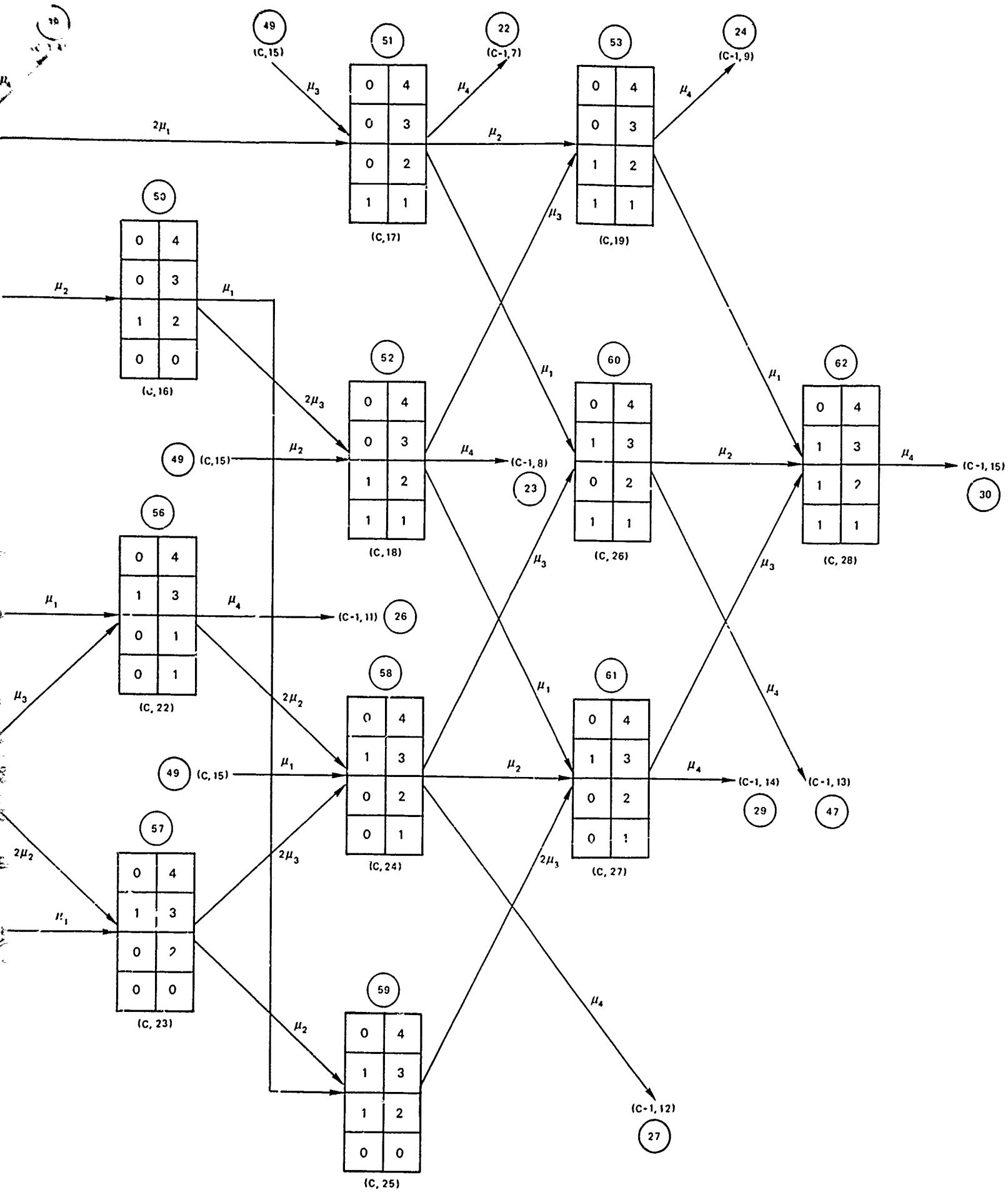


FIGURE III-8b

SERVICE SYSTEM STATES
WITH FOUR CUSTOMERS
PRESENT AND INTERNAL
QUEUES

service system states that occur when there is contention for service stations and interim queues exist preceding states 2, 3, and/or 4. Transitions between states are represented by arrows with associated probability coefficients. Figure III-8a shows transitions to Figure III-8b states, and Figure III-8b shows the Figure III-8a states from which transitions to Figure III-8b states are made. Transitions corresponding to customers leaving the service system are shown only for stage 4 service completions. The concise state notation (C, X) is used where $C = c_1$, and X is the state number assignment. When a customer leaves the system a transition of the form $(C, X) \rightarrow (C-1, Y)$ occurs. When $C = 4$ the transition is from state X in Figure III-8a or III-8b to some state Y in Figure III-7. When $C > 4$, however, the transition is to some state Y in Figure III-8a, and the number of customers being served remains the same, but the arrival queue is reduced by one.

C. Flow Graphs for Statistical Factor Calculation

1. Steady State Model

Calculation of the steady state (equilibrium) probabilities for the occupancies of various states of the system described by Figures III-4 through III-8 is required for subsequent analyses. Recall that the various states have to do with how many contacts are in each part of the system. The steady state probabilities are then mathematical statements of the likelihoods of varying degrees of load and overload on the ASW unit functions.

Flow graph techniques are employed to calculate these probabilities. Flow graphs can be thought of as networks of nodes and arcs through which "probabilities" of certain events "flow." In a sense the nodes and arcs affect these flows in a manner analogous to the way contacts "flow" through the element queuing model. By properly selecting the transmission

characteristics of the network, the transmission through the network is made to be numerically equal to the desired probability.^{8,9} Motivation for and development of flow graph techniques and computational methods is provided in Appendix A.

The steady state (equilibrium) probabilities for the various states of the queuing model described in the previous section are obtainable from the flow graph shown in matrix form in Figure III-9, which combines the diagrams, with certain modifications, of Figures III-4 through III-8. One of the modifications involves the diagram of Figure III-8. To form a complete flow graph the diagram of Figure III-8 must be replicated as often as is necessary to account for various possible numbers of contacts in the first queue (ahead of detection). Theoretically, there could be an infinite number of contacts in queue, but practical solutions are attainable only with a finitely terminated flow graph. There are two practical ways to terminate the flow graph. One is to assume a large but finite population of contacts and simply ignore all arrivals after a certain number of contacts enter the queue. This case is illustrated in Figure III-10. Since, under quite liberal conditions, the probabilities associated with very large queue occupancies are small, the assumption of a finite population has a very minor effect on the system's queuing statistics. The second method of terminating the flow graph is to "fold" the infinite flow graph as illustrated in Figure III-11. A mathematical artifice is employed, and no assumptions of finiteness are necessitated. Some complexity is added to the computation.

The flow graphs in Figures III-9 to III-11 have the same graphical structure (i.e., node-arc incidence properties) as the transition rate diagrams in Figures III-4 to III-8, but the arc flow factors in Figures III-9 to III-11 are the p_{ij} 's of the corresponding Markov process, as explained in Appendix A.

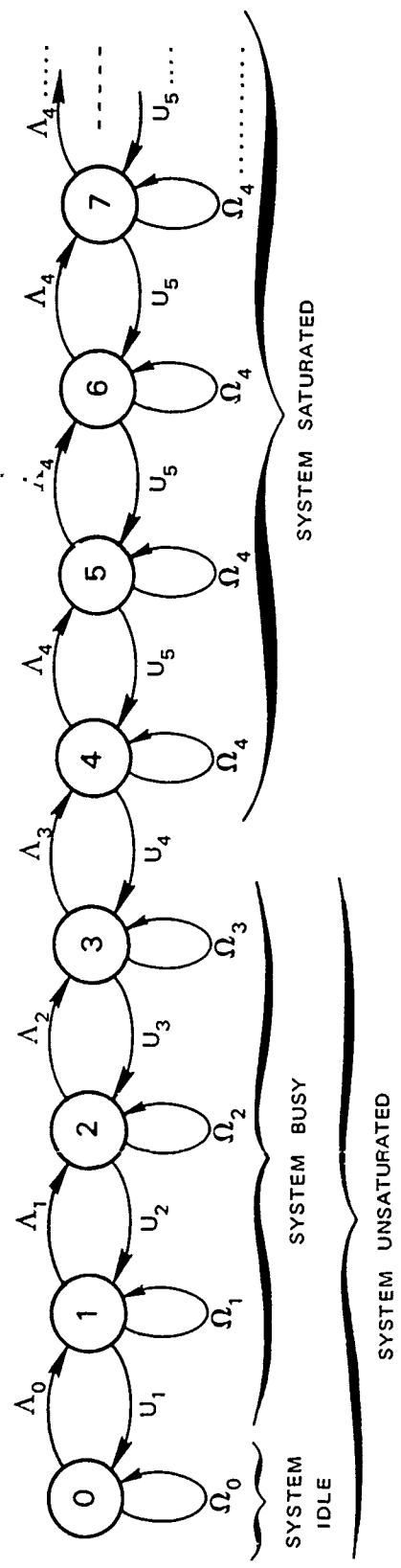


FIGURE III-9 STEADY-STATE MATRIX FLOW GRAPH FOR THE UNIT MODEL

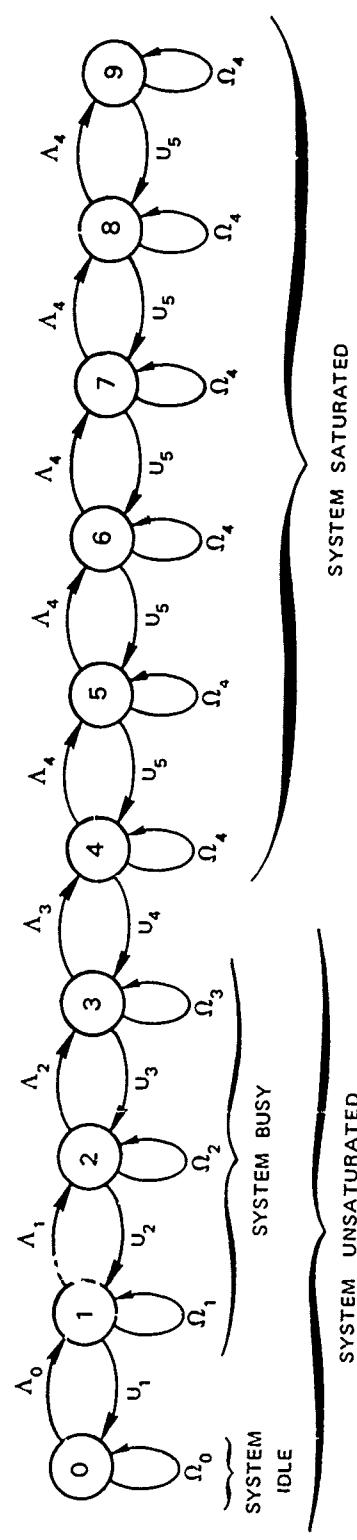


FIGURE III-10 TRUNCATED MATRIX FLOW GRAPH

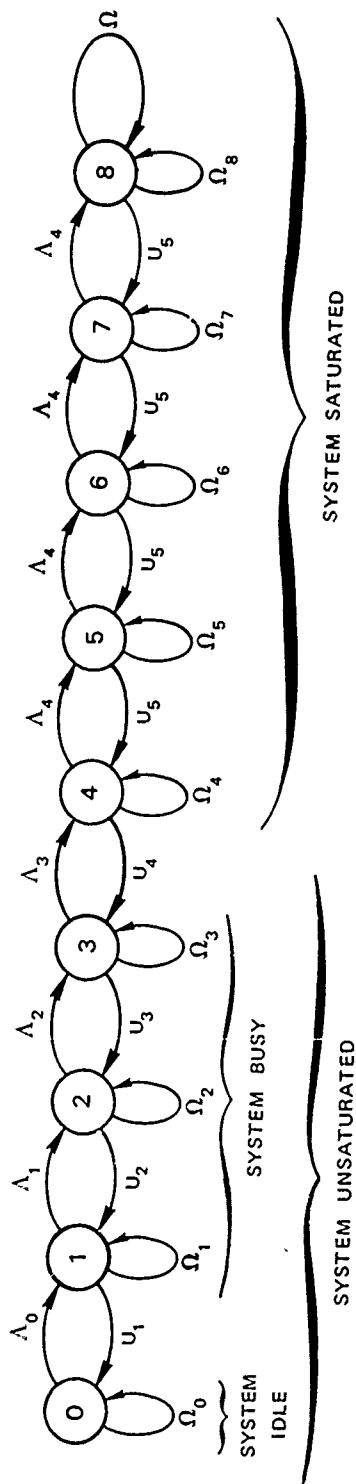


FIGURE III-11 "FOLDED" INFINITE CHAIN IN STEADY-STATE FLOW GRAPH

The vector nodes in Figures III-9 to III-11 correspond to the blocks of states having 0, 1, 2, ... contacts in the system. The 0 state is a scalar state only. The matrices Λ_i , U_i , and Ω_i contain transition probabilities of the form

$$p_{ij} = \frac{\mu_3}{\lambda + \mu_1 + \mu_2 + \mu_3 + \mu_4}, \text{ etc., } \quad (\text{III-C-1})$$

corresponding to the appropriate service and arrival transitions in Figures III-4 to III-8. Note that these matrices repeat after a certain stage. Their internal structure, using the node labels of Figures III-4 to III-8, is obtainable by expressing the node to node incidence relations for the graphs in these figures. This was done for the computer programs described in Section III-D.

The "folded" infinite chain in Figure III-11 requires some explanation. In flow graph applications there is often an infinite chain of repeating identical subgraphs. For example, the simple "birth-death" type of process, i.e., a Poisson queue, has a flow graph as in Figure III-9, but with all Λ 's, U 's, and Ω 's being scalars and given by the following expressions:

$$\begin{aligned} \Omega_0 &= \frac{\mu}{\lambda + \mu}, \quad \Omega_i = 0 \quad i = 1, 2, \dots \\ \Lambda_i &= \frac{\mu}{\lambda + \mu}, \quad i = 0, 1, 2, \dots \quad (\text{III-C-2}) \\ U_i &= \frac{\mu}{\lambda + \mu}, \quad i = 1, 2, \dots \end{aligned}$$

"folding over" amounts to equating the effect of the remainder of all nodes and arcs in an infinite flow graph past a certain point to a single self-

loop with factor Ω . To determine the value of Ω in this scalar case, we calculate it from the relation between Ω and the other problem parameters in Figure III-12, which expresses the fact that the infinite "tail" with

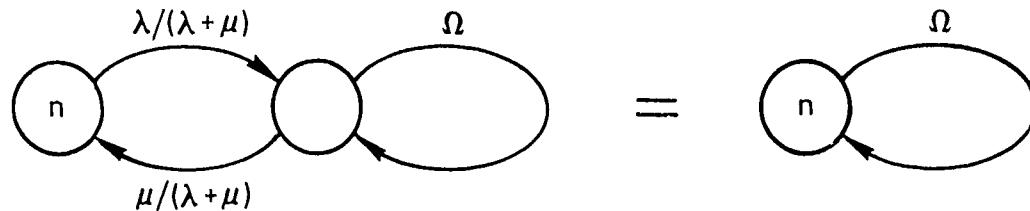


FIGURE III-12 CALCULATION OF INFINITE "TAIL" VALUE IN A FLOW GRAPH

one more repeating item attached is the same as the "tail" itself. But the flow graph on the left side of Figure III-12 can be reduced (by standard flow graph reduction methods⁸) to a self-loop having a transmission equal to:

$$\Omega' = \frac{\lambda\mu}{(1-\Omega)(\lambda+\mu)} . \quad (\text{III-C-3})$$

Thus we have

$$\Omega = \Omega' = \frac{\lambda\mu}{(1-\Omega)(\lambda+\mu)} , \quad (\text{III-C-4})$$

which yields

$$\Omega = \frac{1 \pm \sqrt{1 - 4\lambda\mu/(\lambda+\mu)^2}}{2} = \frac{\lambda + \mu \pm |\lambda - \mu|}{2(\lambda+\mu)} = \begin{cases} \frac{\lambda}{\lambda+\mu} \\ \frac{\mu}{\lambda+\mu} \end{cases} . \quad (\text{III-C-5})$$

This results in two possible solutions for Ω in the interval $(0, 1)$. It will be observed that taking $\Omega = \lambda/(\lambda+\mu)$ leads to a consistent set of

steady state equations, i.e., the finite P matrix arising from the truncated network of Figure III-11 (with scalar nodes and arcs for a birth-death process) is stochastic, whereas taking $\Omega = \mu/(\lambda+\mu)$ leads to an inconsistent set of equations. It can also be shown that the birth-death process is stable only when $\lambda < \mu$.

In the general cast of Figure III-11, the "folding" process yields the matrix equation:

$$\Omega = \Omega_4 + U_5(I-\Omega)^{-1} A_4 . \quad (\text{III-C-6})$$

When A_4 is a diagonal matrix with a uniform constant λ running down the diagonal, as is the case in the single unit model, this simplifies to the matrix quadratic equation

$$\Omega^2 - (I+\Omega_4)\Omega + \Omega_4 + \lambda U_5 = 0 , \quad (\text{III-C-7})$$

for which it is possible to iterate towards a solution matrix Ω . The convergence of certain iterative procedures to solve the above equation was considered, but this approach has not been used to date.

One potential disadvantage of "folding" may appear to be that a full matrix Ω is created out of the very sparse matrices Ω_4 , U_5 , and A_4 , because of the inversion process (or, equivalently, solving of the quadratic equation). Thus, whereas at most five multiplications per equation are required (as shown in Section III-D) per iteration in solving for the steady state probabilities, the greater number of 28 operations per equation would be required using the "folded" Ω . However, using Ω allows one to drop all states of the flow graph past some desired truncation point without losing accuracy in the probabilities calculated for the flow graph up to that point. Truncating without folding, as in

Figure III-10, requires going very far out on the flow graph so that any remaining neglected state probabilities would be negligible. This "folding" approach may be considered in future applications but, for present purposes, it has been found that adequate accuracy with acceptable computation requirements is obtained with the simpler truncation method.

2. System Unsaturated Time

As explained in Appendix A, Section 6, probability distributions for the time a Markov or semi-Markov process spends in any given subset of its states is obtainable, in terms of Laplace transforms, by suitably modifying the semi-Markov type of flow graph for the process. The present system is "unsaturated," by definition, when not all the sonar servers are occupied, and therefore new detections are possible. This means that the appropriate subset S of nodes in the flow graph to be considered should be all nodes corresponding to 0, 1, 2, or 3 contacts in the system. The desired probability that the system spends at least time t in states 0, 1, 2, 3, is therefore the inverse Laplace transform of the transmission from input to output for a suitably modified flow graph, as explained in Appendix A. The appropriate flow graph for computing this for unsaturated time is given in Figure III-13 in matrix form. For the sake of simplicity the complete detailed flow graph for Figure III-13 is not shown. The details of the particular subgraphs used for 1, 2, 3, and 4 or more contacts

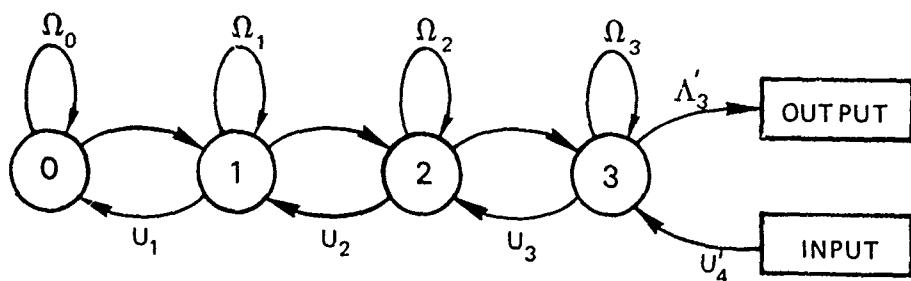


FIGURE III-13 UNSATURATED TIME MATRIX FLOW GRAPH

in the system are given in Figures III-4 to III-8 in structural form, that is, with transition rates along the arcs and not the Laplace transforms needed in the semi-Markov flow graph. Figures A-8a, A-9a, and A-9b in Appendix A illustrate the form of the Laplace transforms.

The matrix U'_4 in Figure III-13 is obtained by replacing each nonzero Laplace transform entry in U_4 by an appropriate steady state probability and "compressing" this new matrix down to a column vector. The matrix Λ'_3 is just Λ_3 "compressed" to a row vector. This "compressing" is all that is necessary since every substate for n contacts in the system of the model connects to at most one substate for $n-1$ or $n+1$ contacts in the system. If there had been more than one possible transition to $n-1$ or $n+1$ contacts, it would be necessary to sum all transition terms in or out of a node in forming Λ'_3 and U'_4 .

Although in principle the inverse Laplace transform of the transmission from input to output for Figure III-13 can be calculated, it was deemed simpler to solve the associated linear system of differential equations. This is discussed further in Sections III-D and III-E.

If it is desired to calculate the probability density function for unsaturated time, it suffices to omit the $1/s$ factors on the output arcs of Figure III-13 (as pointed out in general in Appendix A).

3. System Saturated Time

In a manner similar to that employed to determine the distribution of system unsaturated time, an appropriate modification of the semi-Markov flow graph of the model is formed for saturated time. In this case only states corresponding to 4 or more contacts in the system are used. Unfortunately, in this case we have the infinite flow graph of Figure III-14. Here Λ''_3 and U''_4 play reverse roles of their counterparts Λ'_3 and U'_4 in Figure III-13. That is, Λ''_3 now is a column vector of normalized steady state probabilities, and U''_4 now is a row vector having

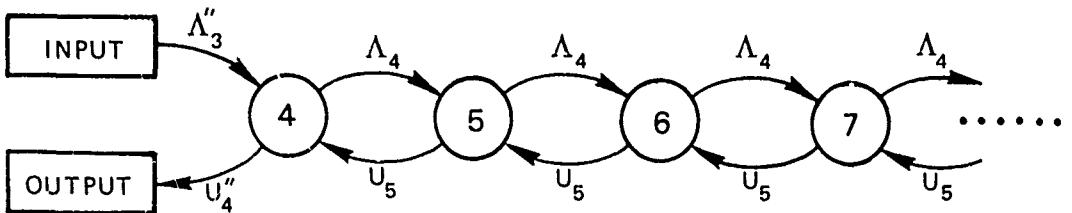


FIGURE III-14 SATURATED TIME MATRIX FLOW GRAPH

components $1/(s(s+\mu_4))$ in certain columns corresponding to contacts leaving the system.

Since the above flow graph is infinite, there are various ways to proceed. One way is to "fold" the flow graph and invert the resultant Laplace transform. Unfortunately, folding in this case leads to a rather complicated irrational Laplace transform, i.e., involving numerous square roots of many rational terms. This leads to complicated combinations of Bessel functions, in fact, an infinite series of Bessel functions, for the inverse transform. Thus only an approximate solution can be obtained, even though the exact Laplace transform is known. Howard⁶ has used this approach successfully on relatively small problems. It may be more efficient, from the numerical computation standpoint, to truncate the flow graph at some suitable stage and solve a finite system of constant coefficient linear differential equations. An alternative would be not to truncate but formulate the equivalent time dependent coefficient linear equations or integral equations that arise in situations like this, and solve these by numerical integration. For ease of implementation, truncation of the flow graph was employed to derive the illustrative results.

4. Waiting Time

To calculate a distribution for the time a contact must wait in the first queue (detection queue), it is necessary to go beyond just taking subsets of the total set of states of the system. The progress

of a specific contact must be considered and not just the state of the total system. To the extent that the waiting of a contact involves the system's having 5 or more contacts in it, the subset of states with 5 or more contacts is in fact going to be considered, but the transition rates between these states have to be modified to reflect the service priority that is established. For this purpose all stimuli awaiting detection are assumed to be equally likely to be detected by the sonar. If a first-come-first-served policy had been chosen as a service priority for the waiting stimuli, as is common in many queuing models, the mathematics of the waiting time model would be simpler--at the expense of considerably more complexity in the time-in-system model of the next section. For this reason random selection of stimuli for detection was assumed.

On the basis of random selection from the first queue, a sonar stimulus has a probability of $1/n$ for being detected. The probability that some other sonar stimulus is detected is $(n-1)/n$, given that the system is in the process of detecting. The correct flow graph modification to represent this phenomenon is given in Figure III-15. The matrices U'_5 in Figure III-15 are copies of the matrix U_5 compressed to a row vector (and multiplied by $1/s$ if the cumulative distribution is desired). The matrix U_5 is understood to contain the necessary semi-Markov flow graph terms of the type $\mu_4/(s+\lambda+\dots+\mu_4)$, just as in the previous two cases. The matrices Λ'_5 , Λ'_6 , ... are structurally all identical, i.e., copies of the Λ_4 compressed to a column vector and having nonzero entries only at the same rows as Λ_4 . These entries are now required to be the steady state probabilities for the corresponding substates for n in the system, etc., instead of the terms $\lambda/(s+\lambda+\dots)$ used in forming Λ_4 .

As in the case of system saturated time, the flow graph for waiting time is infinite, and one can either truncate it or attempt to fold it over into an equivalent finite flow graph. Folding has not been

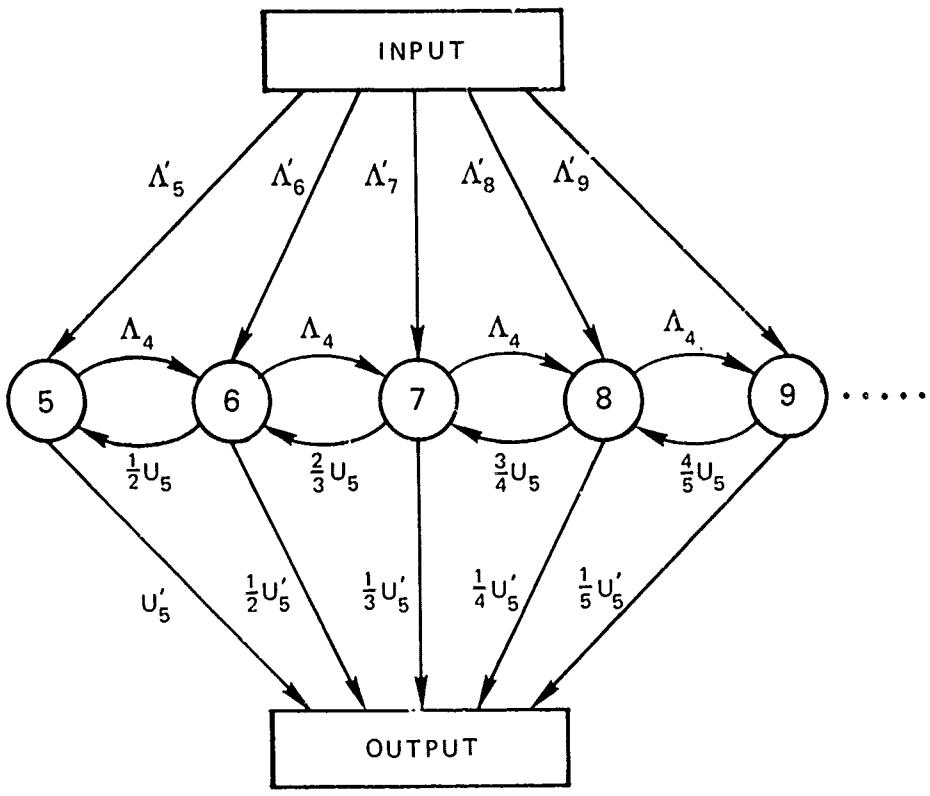


FIGURE III-15 WAITING TIME FLOW GRAPH FOR RANDOM SELECTION FROM STIMULUS QUEUE

fully investigated in this case thus far. In the case of the first-come-first-served priority policy, for which the flow graph of Figure III-16 was obtained, folding is facilitated by the ability to establish first the equivalent repetitive flow graph of Figure III-17. A similar development for the random priority policy was investigated, but this approach was deemed unlikely to lead to more efficient numerical solution. As with system saturated time, truncation of the flow graph was employed.

5. Time In System

The most complex flow graph for the unit model arises in the case of calculating the probability distribution for time in the system spent by any one specified contact. This is a useful measure from the

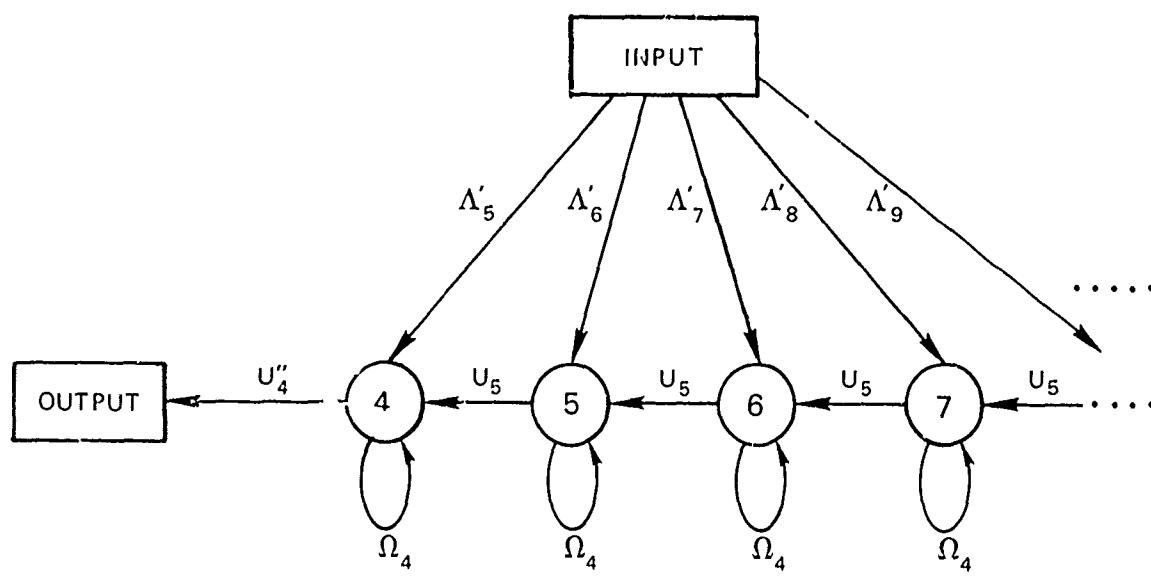


FIGURE III-16 WAITING TIME FLOW GRAPH FOR FIRST-COME-FIRST-SERVE POLICY AT STIMULUS QUEUE

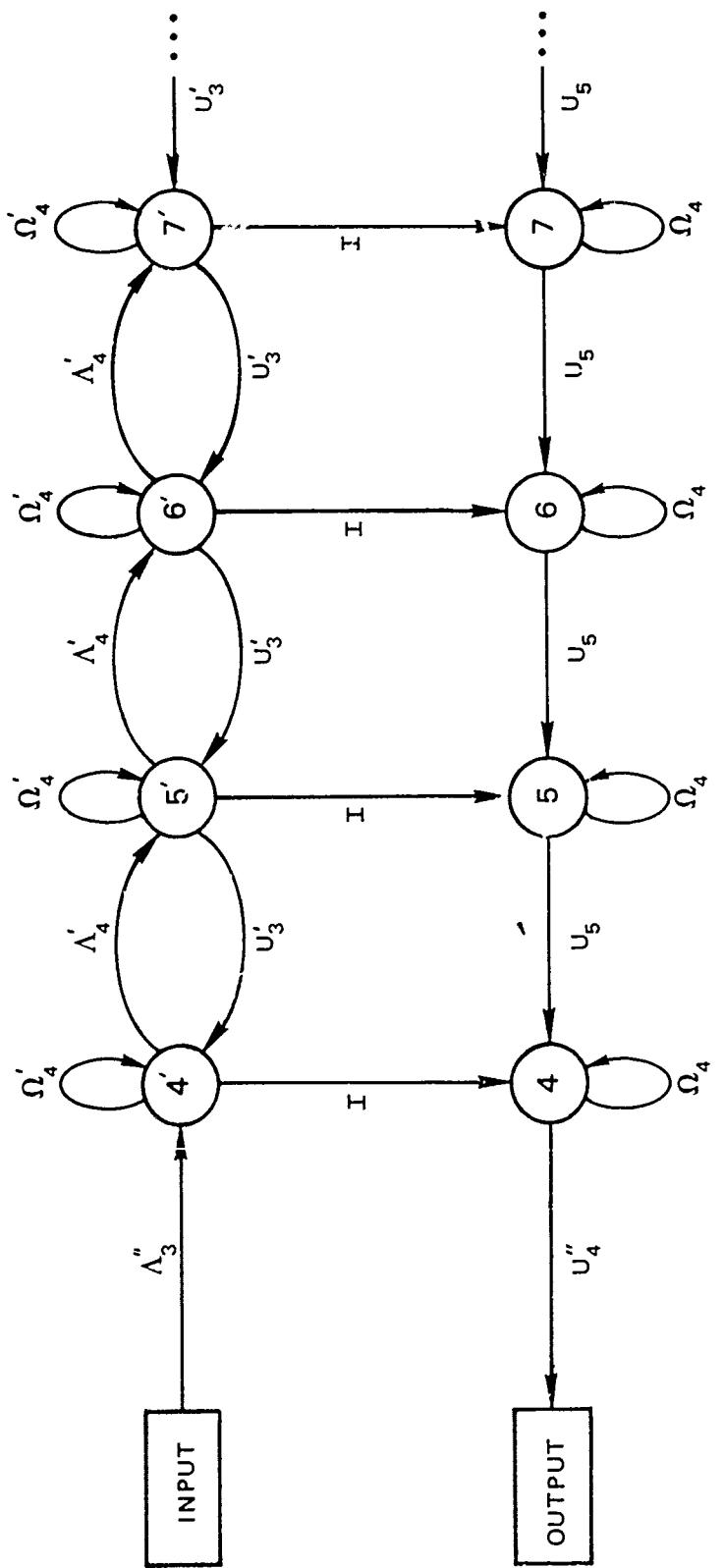


FIGURE III-17 REPETITIVE WAITING TIME FLOW GRAPH FOR FIRST-COME-FIRST-SERVE POLICY

point of view of determining whether the ASW platform is capable of dealing with the contact prior to its approaching the platform "too close for comfort." The other measures above are more related to how well the ASW platform functions, where any bottlenecks might be, etc. They are also useful in a more involved model for target/ASW platform interaction, which is the subject of the other principal area of investigation on this project, viz., the force effectiveness model development.

A suitable flow graph for computing the Laplace transform of the probability that a given contact spends time t or less in the system may be derived as follows. As in the case of waiting time, the progress of a specific contact must be taken into consideration. Essentially, a Markov process is created with trapping states placed in such a way that, no matter what is happening to other contacts, process termination occurs only at states that correspond to the specified contact exiting the system. This is complicated because contacts can pass each other in the model. In order to represent all possible states of the system relative to what the specified contact is doing, the dimension of the states must be compounded. That is, for every actual state of the model we must consider as many alternatives as there are possible positions of the specified contact allowed by that state. The highest number of alternatives that will ever appear is 4, when there are no contacts in the primary (detection) queue, since there are at most 4 places (out of the less than or equal to 4 contact positions) that the specified contact could be. When contacts queue up ahead of the detection server, the random selection policy eliminates the need to keep track of where the specified contact is in the queue. With a first-come-first-served policy in the detection queue, the situation becomes still more complicated. With random selection, there is the possibility of contacts passing each other both in the first queue and farther on in the system, because of the multiple service channels at three of the four stages of service. Theoretically at least,

an infinite number of contacts that arrive after a given contact could possibly pass it before it gets out of the system, although this fact is primarily of academic interest only. Nevertheless, even for computational purposes, the above considerations must be taken into account or else a loss of accuracy results.

As in the case of steady state calculations, folding is necessary to maintain accuracy in a truncated flow graph when the probability of the system's being in any of the neglected states is not negligible. In the present case folding seems to be even more difficult than in the case of the waiting time distribution.

Returning to the derivation of the flow graph, let us define copies i_a, i_b, \dots of a given state i , using notation of Figures III-4 to III-8 for the description of a state. We define the copy i_a to represent the given contact being in the first available position it could be in for state i , counting from the top down and from the left to the right in Figures III-4 to III-8. Thus, for example, state 52 from Figure III-8b will have the four copies in Figure III-18. Note that only one contact can be in each of the 3 queues after the first, so no subdivision is necessary to indicate where in the queue a contact lies. The circles in Figure III-18 denote the specified contact's position. Partial details of the resultant transition rate diagrams for 2, 3, and 4 contacts in the

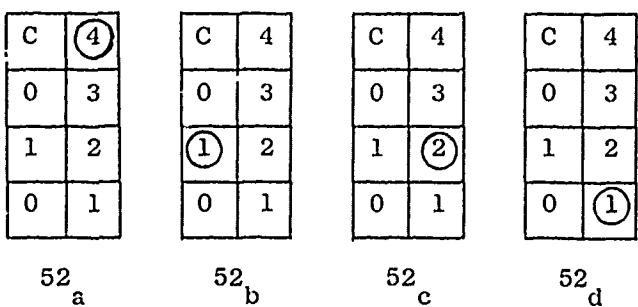


FIGURE III-18 THE FOUR COPIES OF STATE 52 OF FIGURE III-8b WITH POSITION OF SPECIFIED CONTACT DESIGNATED

system are given in Figures III-19 to III-21. As a result, when comparing the time-in-the-system model to other measure of effectiveness models, there is a greatly increased number of states for each number of contacts in the system. The magnitude of this increase is indicated in Table III-1.

Table III-1

NUMBERS OF STATES IN SUBGRAPHS
FOR GIVEN NUMBER OF CONTACTS IN SYSTEM

No. Contacts in System	No. States in Subgraph	
	Time in System Model	Other Models
0	1	1
1	4	4
2	17	10
3	44	19
4 or more	83	28
Totals for up to 4 in system	149	62

The already large number of 149 of states for just up to 4 contacts in the system is only the beginning. Next must be added an infinite array of copies of the subgraph of Figure III-21 to represent the various cases where the singled out contact is in service while particular numbers of other arrivals are in queue, since it matters whether others are in queue and how many are. This is because the emptying or nonemptying of the queue affects the singled out contact's competition for service. Thus, if the particular contact gets into service when there are a great many in queue, this leads to a higher probability of its being passed more times and therefore encountering blocked servers more often than if there were fewer competitors. Finally, the waiting time flow graph must be

appended to the time-in-system flow graph to represent the cases where the contact is in queue along with other contacts. To accomplish this, the waiting time flow graph's output arcs are connected to appropriate points in the time-in-system graph, corresponding to the singled out contact going into service with particular numbers of other contacts still left in the queue. The final result is the matrix flow graph of Figure III-22. Note that there is no zero node since a contact does not spend time in the system with zero contacts in it.

The double circled vector nodes in Figure III-22 represent the full extensions of Figures III-19 to III-21; that is, these nodes correspond to the singled out contact's being in service with a total number of 2, 3, or 4 contacts in the system. The modified transition matrices Λ'_i , U'_i , Ω'_i , and U''_5 reflect the necessary changes that must be made in the structure of the semi-Markov process to allow it to transit between states where the specified contact is competing in service and states of the unmodified Figures III-8a and III-8b are represented by the single circled nodes of the waiting time section of the flow graph. The "prime" notation indicates that these matrices are variations on the original ones. Their detailed structure is rather complicated and is not given here, but is easily derivable once an indexing scheme for the states is established.

D. Computational Approaches and Sample Solutions

1. Steady State (Equilibrium) Distributions

The steady state distribution for the single element model truncated at various numbers of contacts in the system was calculated for several values of arrival rate λ and a fixed set of service rates μ_1 to

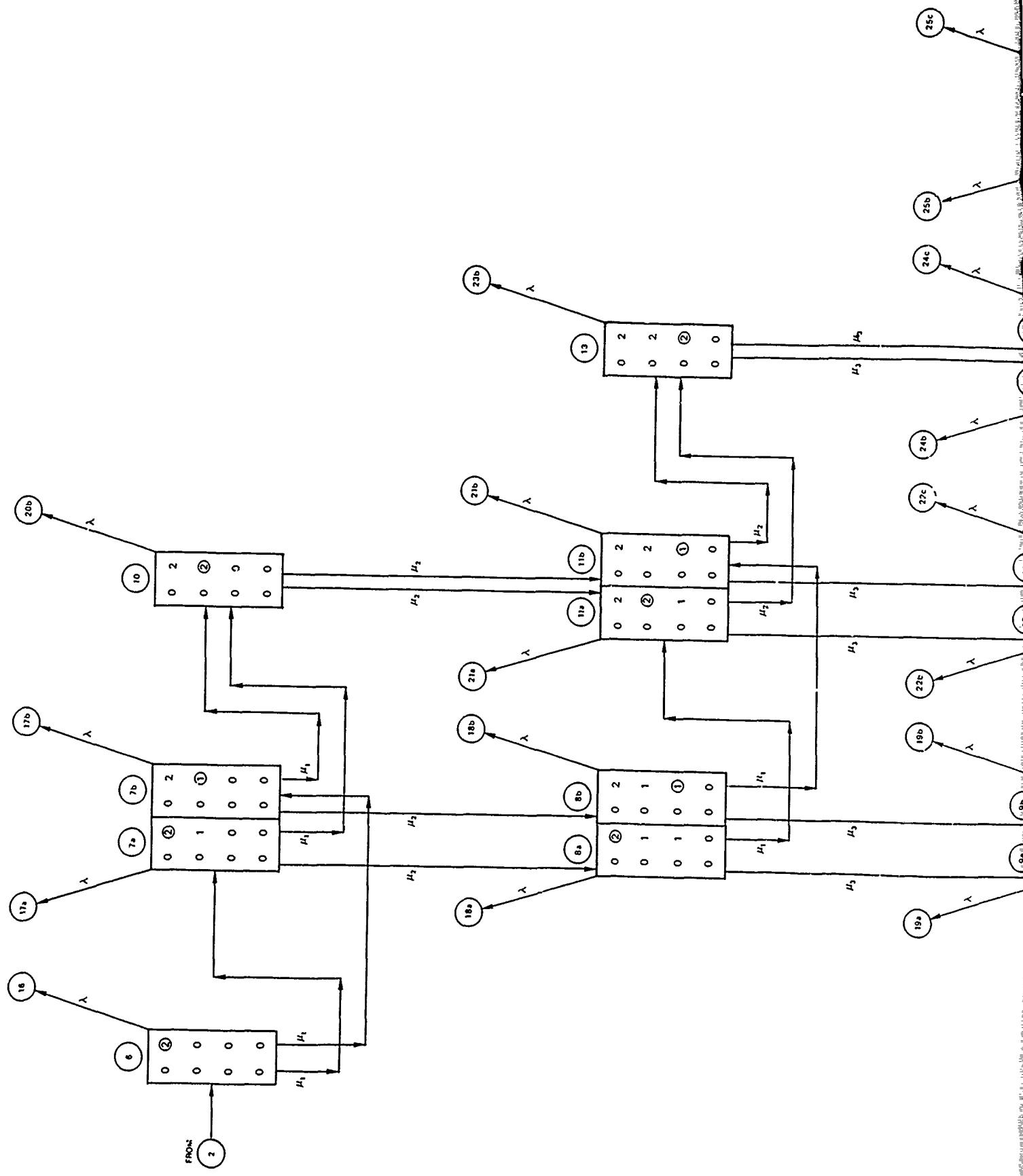
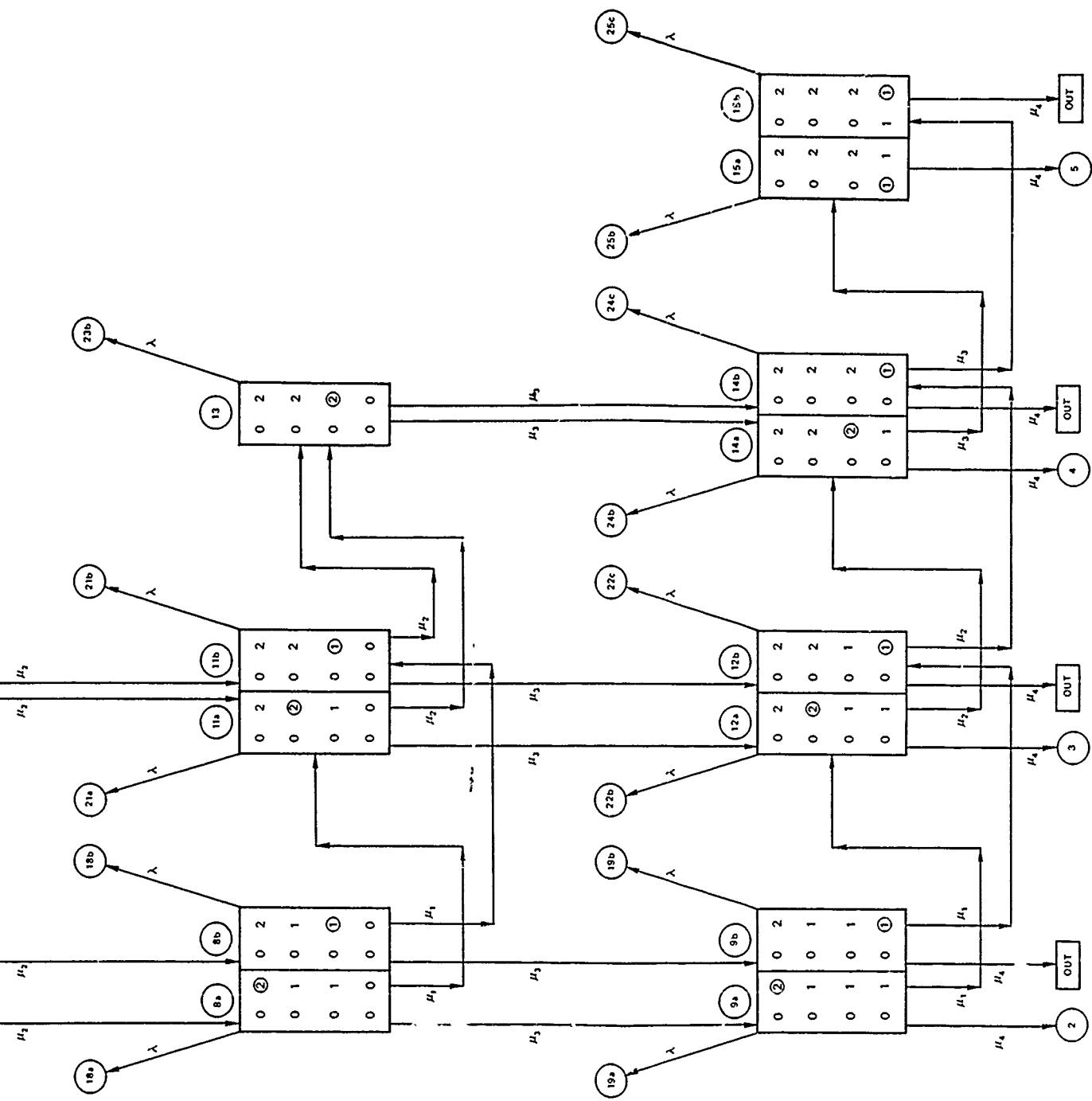
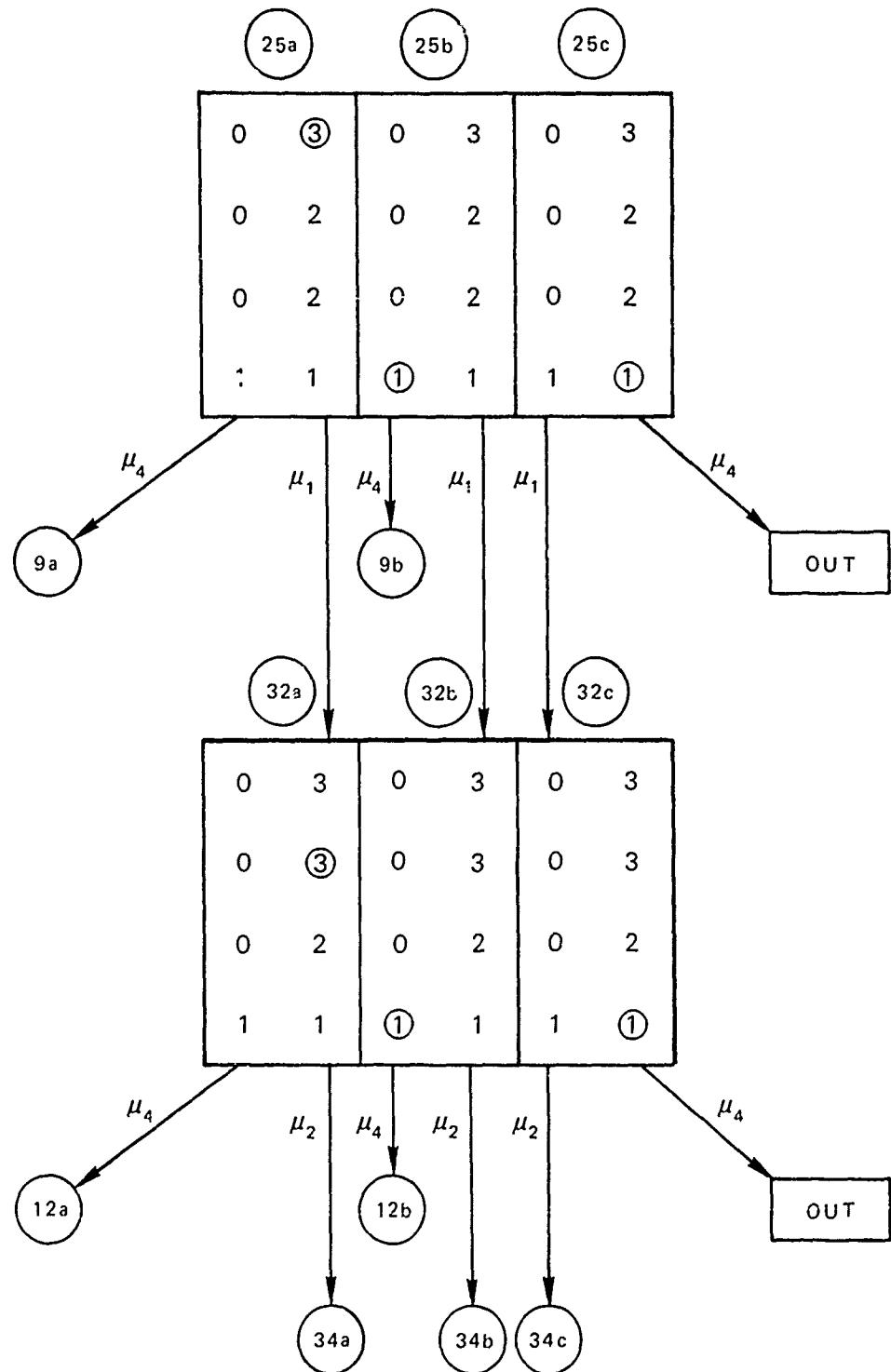


FIGURE III-19 DETAILED DIAGRAM FOR TWO IN SYSTEM WITH POSITION OF SPECIFIED CONTACT DESIGNATED

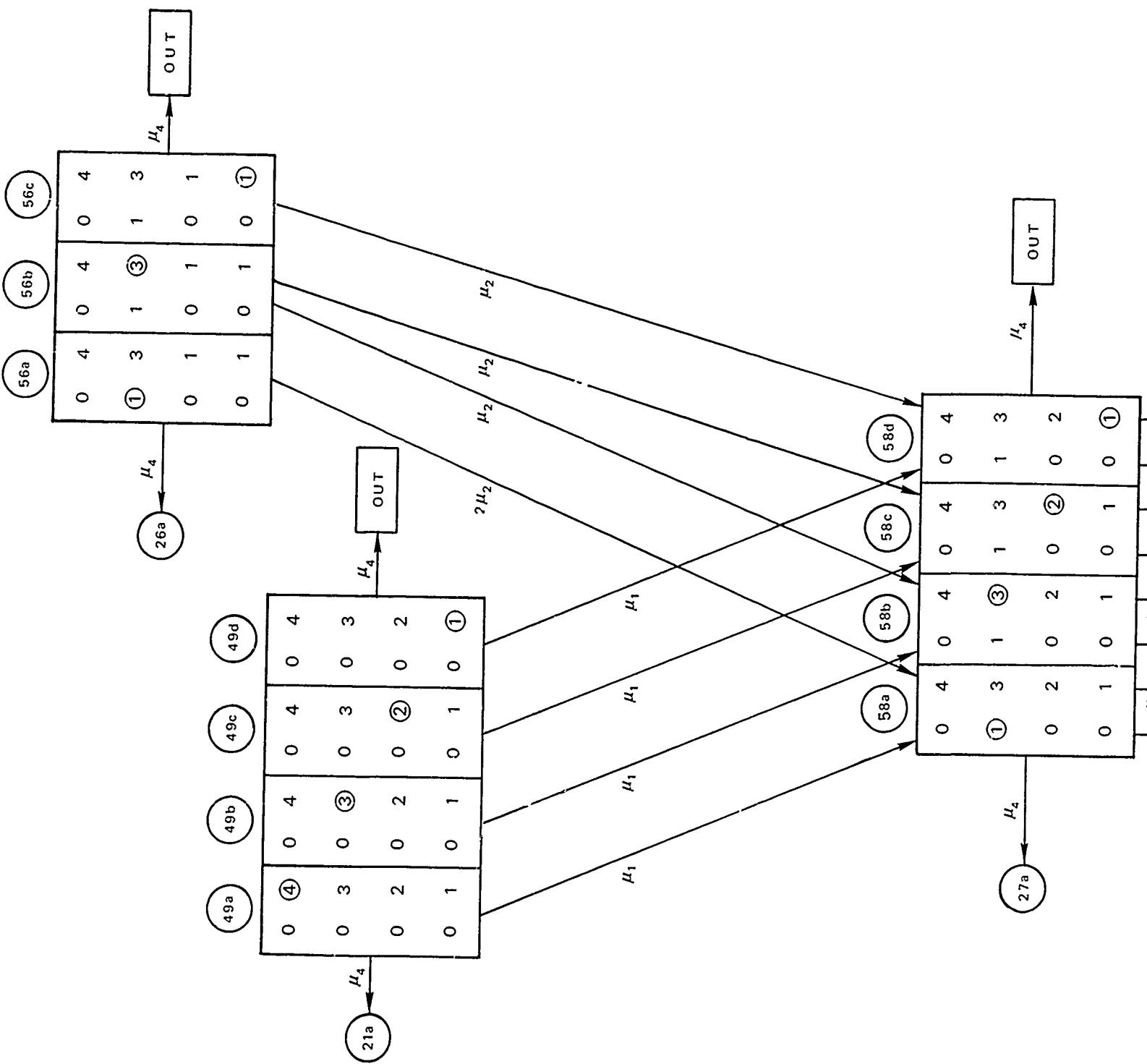


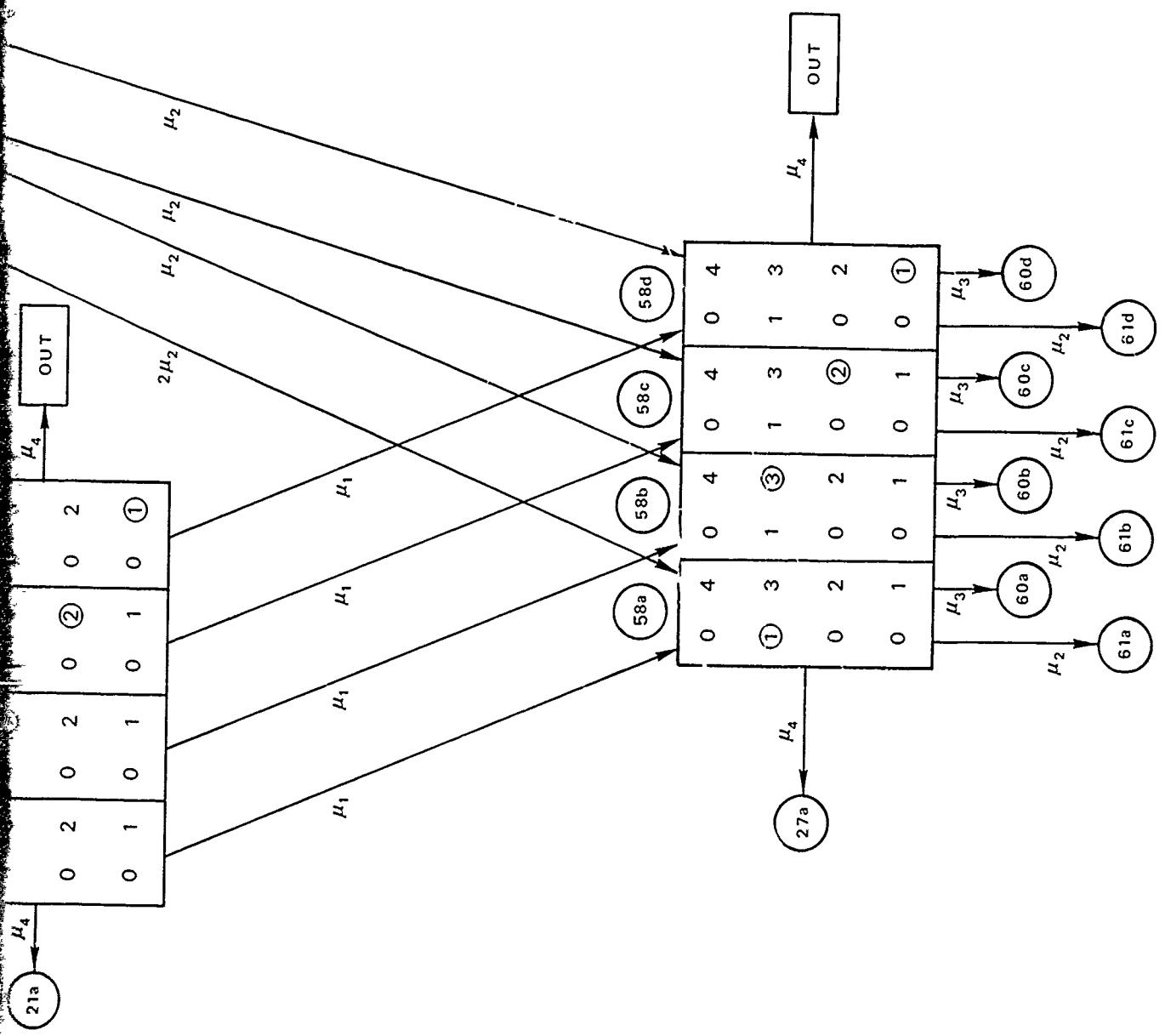
29



NOTE: TOTAL NUMBER OF STATES = 44

FIGURE III-20 PARTIAL DETAILS OF DIAGRAM FOR THREE IN SYSTEM WITH POSITION OF SPECIFIED CONTACT DESIGNATED





NOTE: TOTAL NUMBER OF STATES = 83

FIGURE III-21
PARTIAL DETAILS OF DIAGRAM
FOR FOUR IN SYSTEM WITH
POSITION OF SPECIFIED CONTACT
DESIGNATED

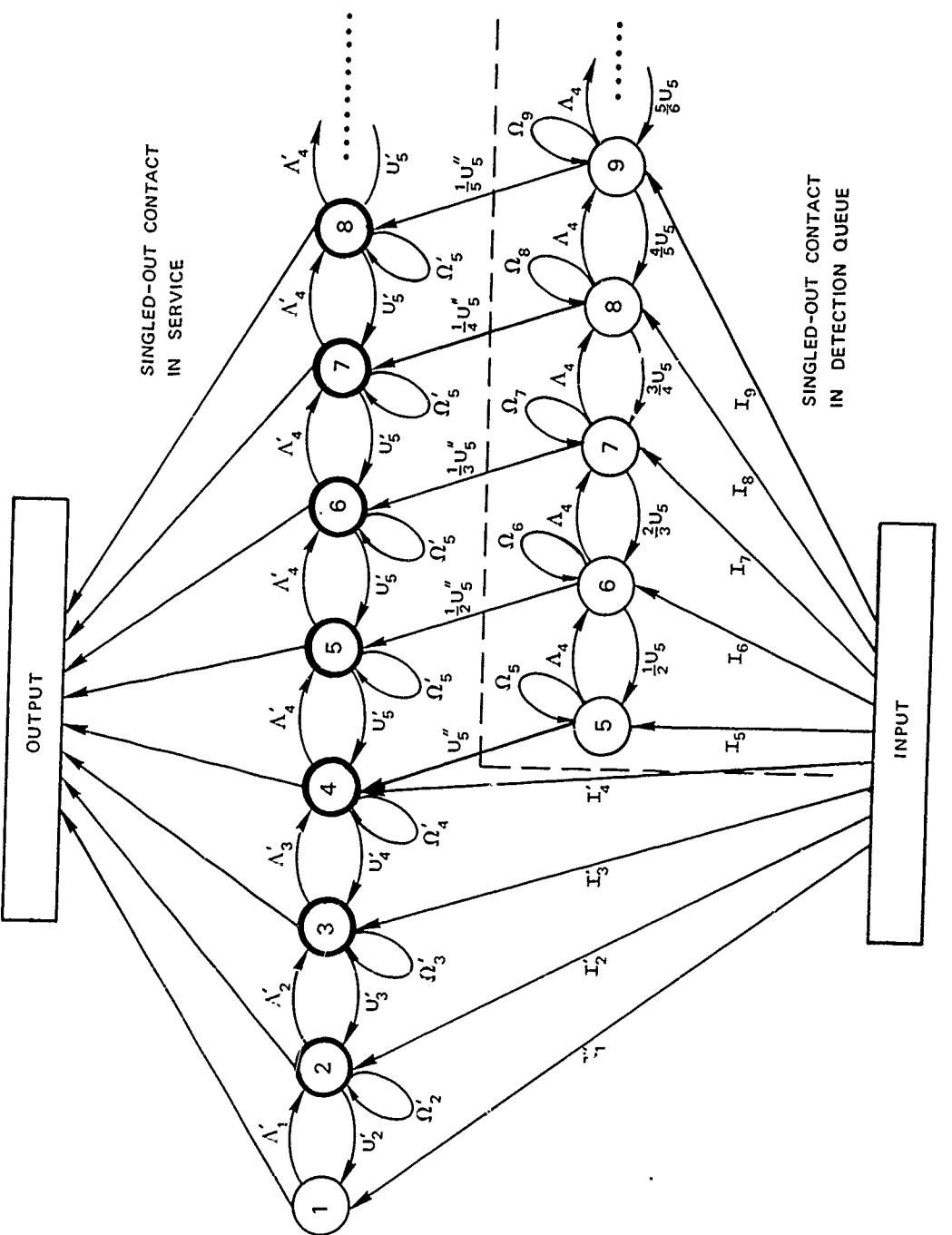


FIGURE III-22 TIME-IN-SYSTEM FLOW GRAPH

μ_4 . The resulting distributions are presented in Figures III-23 and III-24. Figure III-23 clearly shows the need to include the possibility of a large number of contacts in the system before acceptable accuracy is obtained. As mentioned earlier, this could be avoided by folding, but application of the folding approach was not completed at the time of writing of this report.

The interpretation of Figures III-23 and III-24 in ASW terms is as follows (the parameter values chosen for these illustrative computations are hypothetical). Figure III-23 shows the likelihood of finding a particular number of contacts in process in the system, given particular average contact arrival and service times. Thus, for an average rate of 5 contacts per hour with (fictitious) detection, localization, attack, and kill times (mean values) of 1 minute, 4 minutes, 12 minutes, and 5 minutes, respectively, the most likely number of contacts in the system at one time is 3. The probability of a peak load of more than 5 is about 1/3. More than 5 in the system corresponds to 2 or more contacts "in queue" awaiting detection--an expression of a degree of overloading of the ASW system unit. The several curves in this figure are the several approximations to the exact probability curve. Figure III-24 shows how system load (number of contacts in the system) varies with increasing contact arrival rate.

The method of solution is an iterative one, as opposed to direct solution of the linear equations for the steady state probabilities. The reason for this is that the direct solution would require use of a very large matrix in any of the standard available computer software packages for solving linear equations. This matrix would have a vast majority of

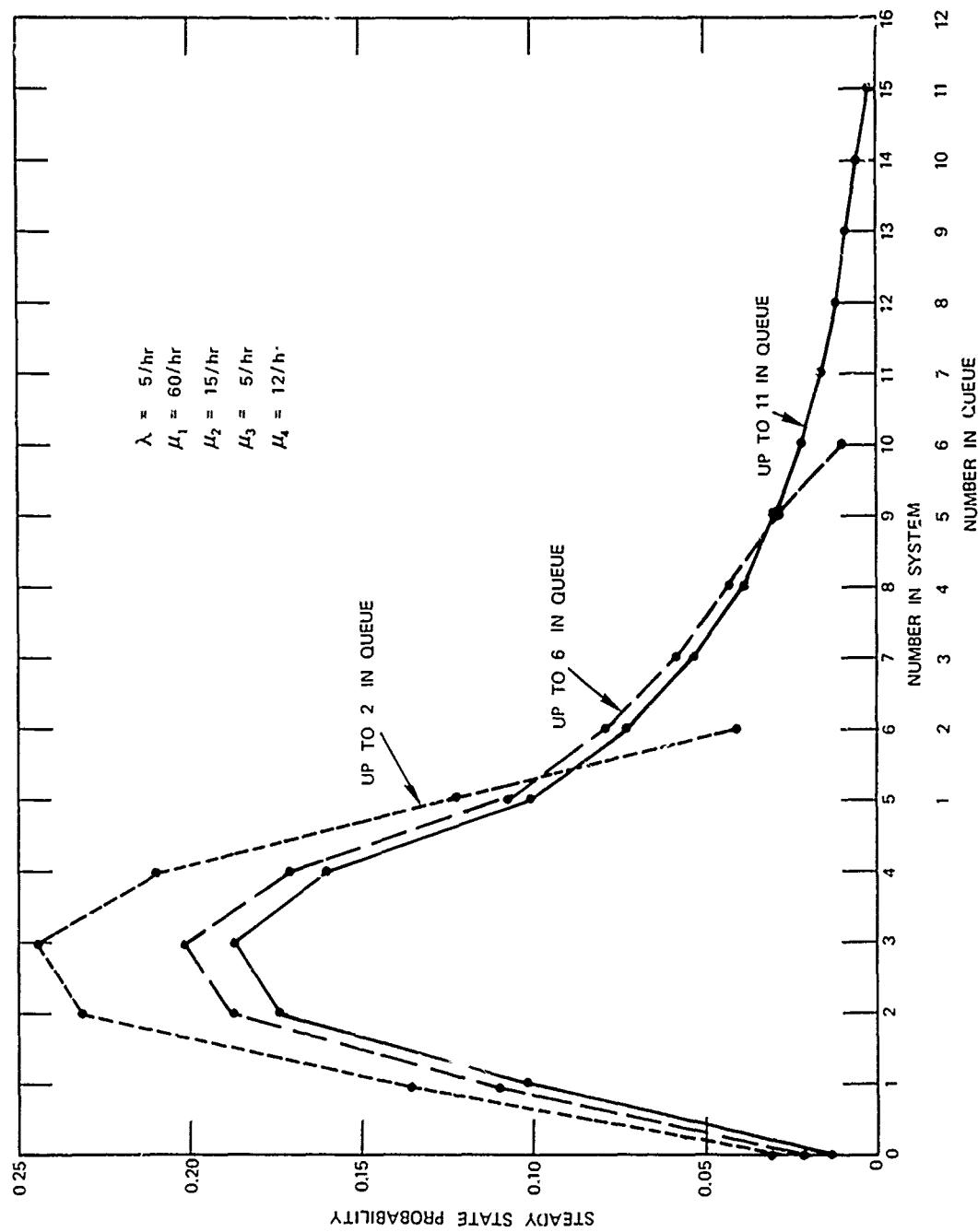


FIGURE III-23 STEADY-STATE PROBABILITY DENSITIES FOR VARIOUS TRUNCATION LEVELS OF STIMULUS QUEUE

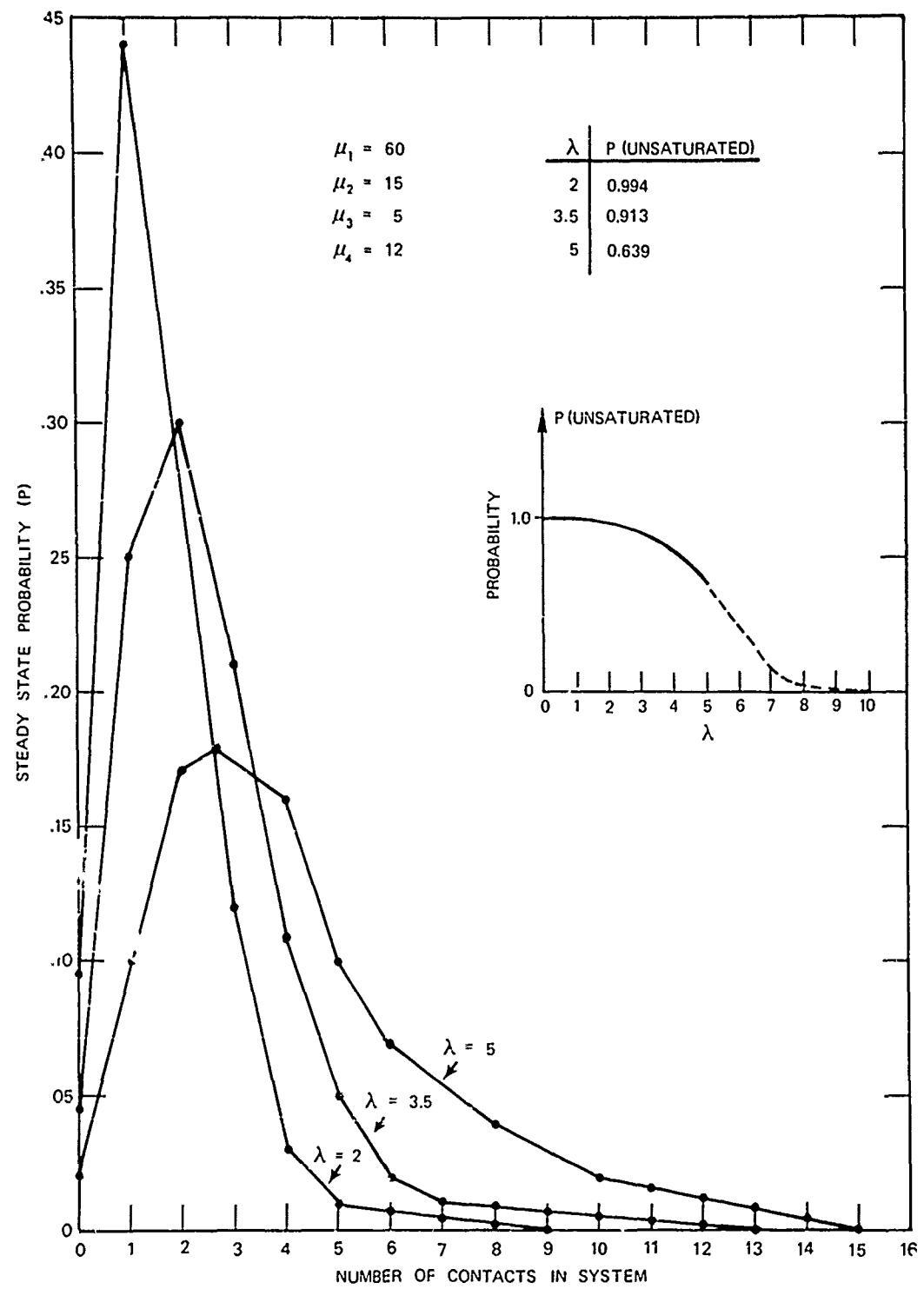


FIGURE III-24 STEADY-STATE PROBABILITY DENSITIES
USING VARIOUS VALUES OF INPUT RATE

zeroes in it, thus rendering the solution extremely inefficient. Therefore, a special algorithm was written for this problem, making use of the fact that each row of the matrix P has at most five nonzero entries, and also making use of the special block structure of the matrix. The method of solution used was the following iteration

$$\lambda_i p_i^{(n+1)} = \sum_{j=1} p_{ji} p_j^{(n)} \lambda_j ,$$

where $\{p_1^{(n)}, \dots, p_N^{(n)}\}$ is the n -th trial set of probabilities. After each iteration, the solution vector is normalized to ensure that numerical errors do not gradually accumulate, which would cause the sum of the probabilities to drift away from a value of unity. The initial trial solution was taken to be the vector $(1/N, \dots, 1/N)$. The above iteration did not converge per se, although it did "converge" to a cyclic set of solutions having period 5. That is, after a certain number of iterations (anywhere from 60 to 500, depending on λ), there began to be a repetition of a cycle of 5 solution vectors. In retrospect, one can easily see why this system should exhibit this behavior. By way of explanation, the matrix P is periodic with period 5, due to the obviously periodic structure of the flow graphs arising from the use of Figures III-4 through III-8 as basic elements. The physical explanation is demonstrated by Table III-2, containing an illustrative sequence of states that begin to repeat every 5 iterations after the 9-th iteration, for up to 3 contacts in the system. For larger numbers in the system the "transient" would be required to propagate for a larger number of iterations before repetition sets in. Clearly, these iterations repeat with a period of 5 because it takes 5 steps to propagate a transfer around each of the loops, and all loops

Table III-2

LIST OF STATES OCCUPIED AS STEADY STATE ITERATIONS
PROCEED STARTING AT STATE NUMBER 1 (0 IN SYSTEM)

		States Occupied					
		(0*)	(1*)	(2*)	(3*)		
	1	1					
	2	2					
	3	3	6				
	4	4	7		16		
	5	5	8	10	17		
	6	1	9	11	18	20	
	7	2	12	13	19	21	26
	8	3	6	14	22	23	27
	9	4	7	15	16	24	28
	10	5	8	10	17	25	30
Iteration No.	11	1	9	11	18	20	32
	12	2	12	13	19	21	26
	13	3	6	14	22	23	27
	14	4	7	15	16	24	28
	15	5	8	10	17	25	30
	16	1	9	11	18	20	32
	17	2	12	13	19	21	26
	18	3	6	14	22	23	27
	19	4	7	15	16	24	28
	20	5	8	10	17	25	30
	21	1	9	11	18	20	32
	22	2	12	13	19	21	26
	23	3	6	14	22	23	27
	24	4	7	15	16	24	28

*Number of contacts in system

have period 5 or multiples thereof in the steady state flow graph of the model, which is a union of copies of Figures III-4 through III-8 with appropriate conversion of the transition rates λ , etc., to normalized p_{ij} 's, etc., (refer to Appendix A and Figure III-25 for explanation).

The interpretation of Table III-2 is as follows. Initially, assume only p_1 has a nonzero value (equal to 1). After two iterations this unit value splits into two quantities summing to unity for states 3 and 6; then these split to yield something for states 4, 7, and 16, etc., (see Figure III-25). Never do any other states receive any nonzero values for their p_i 's except according to the sequence in Table III-2. Thus, no equilibrium can ever be reached because the incoming periodic effects from other nodes always come "in phase" to maintain the periodic behavior of this sequence. Examples of this cycling effect shown in Table III-2 can be traced on Figure III-25.

In the numerical computations of "steady state" solutions the values of the last 5 iterations are averaged because of this periodic behavior. The oscillations remaining after transients died out were about 10% of average in their peak-to-peak value in most components of the probability vector. Exact steady state solutions could be obtained by solving the linear equations by direct inversion of the P matrix (with one row removed, of course, as it is singular and of rank N-1). This was not attempted because of the large size of this matrix when up to 12 contacts in the queue were considered. The actual size of such a matrix would be $N \times N$, where

$$N = 1+4+10+19+28+12 \times 28 = 398 ,$$

and where each term in the sum corresponds to a particular number of contacts in the system (see Table III-2).

Thus far, there is not a complete physical interpretation of this periodic behavior of the solution algorithm, insofar as the lack of existence of a unique steady state is concerned. This is a matter for further investigation. In this connection, it should be recalled that the model is linear, i.e., leads to linear differential equations with periodic solutions of arbitrary amplitude for the "steady state" probabilities as functions of time. Since any practical system is really nonlinear, saturation and self-sustaining oscillations may occur, and there may not, in fact, exist any equilibrium in the true sense, although there may be what are known as stable cycles, i.e., continually repeating sets of states (along some closed trajectories in the state space).

2. Unsaturated, Saturated, and Waiting Time Distributions--Illustrative Results

The probability (cumulative) distribution for system unsaturated time was calculated numerically by integration of the system of differential equations associated with the flow graph of Figure III-13. The method of integration used was basically centered around the so-called "successive overrelaxation iterative method," described in Reference 10, for solution of linear equations. This method is not the most efficient one available but is relatively easy to implement by suitable modification of the steady state computer program. Furthermore, this method required no matrix inversions, and the entire computer procedure is efficiently handled in terms of node-arc incidence of the flow graphs, rather than by storing huge matrices (as in the usual, general approach to differential equation integration). Also, the overrelaxation approach led to easy modification of the unsaturated time program to obtain programs for calculations of saturated time and waiting time distributions. (Due to lack of time, the time-in-the-system calculations were not completed so as to be included in this report.) The mathematical technique is developed in Appendix B. Some illustrative results are presented below.

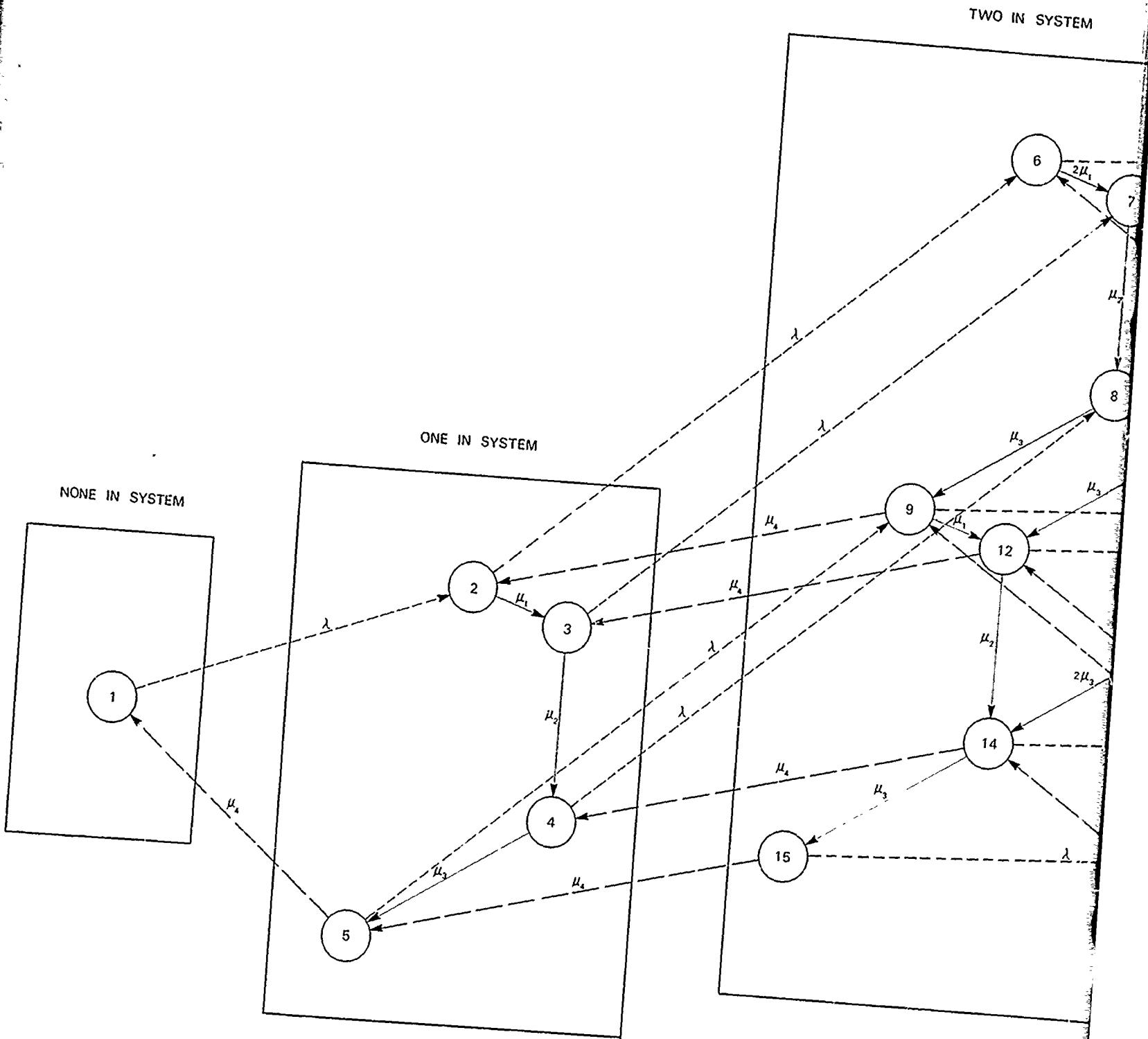
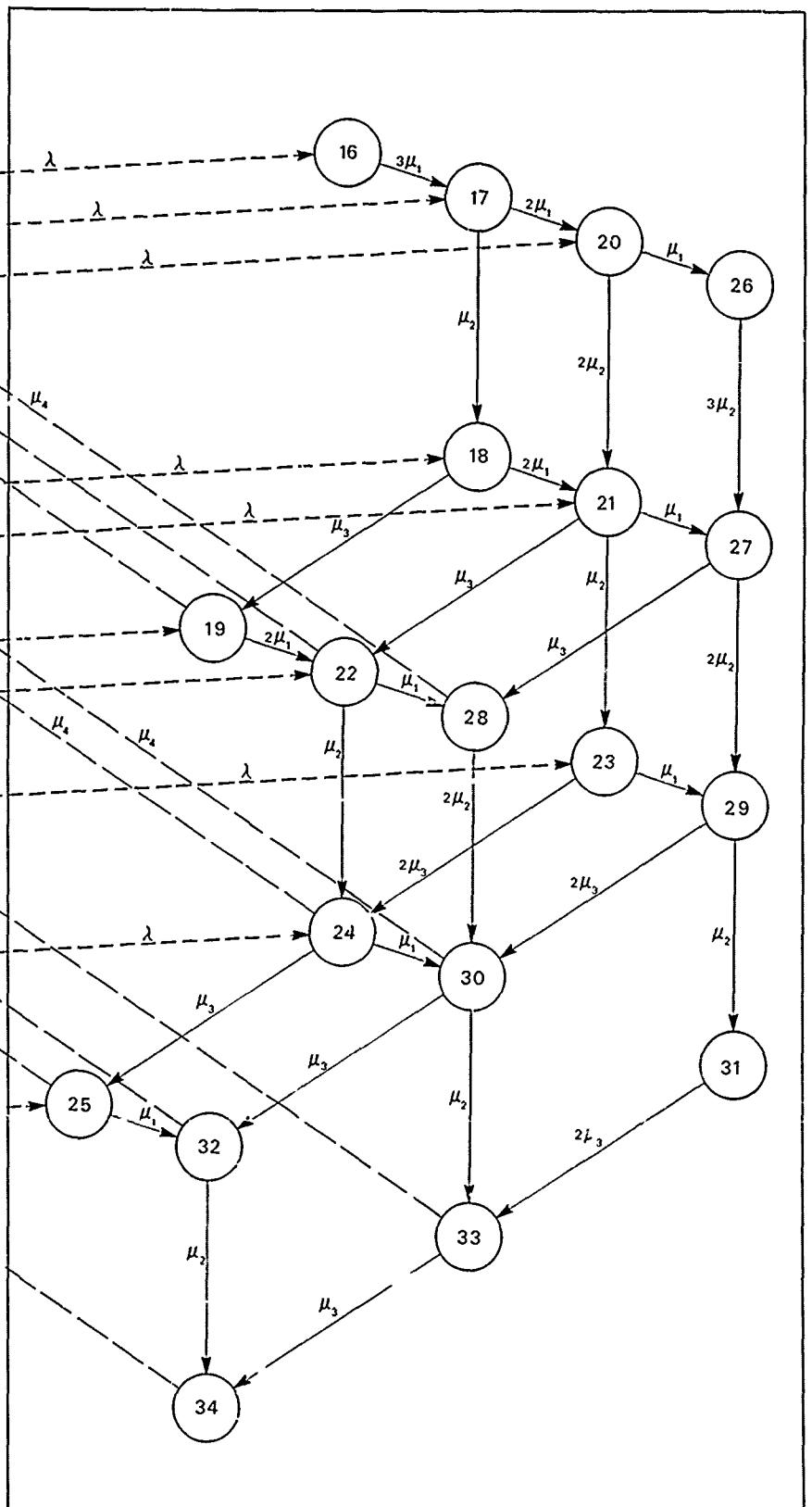
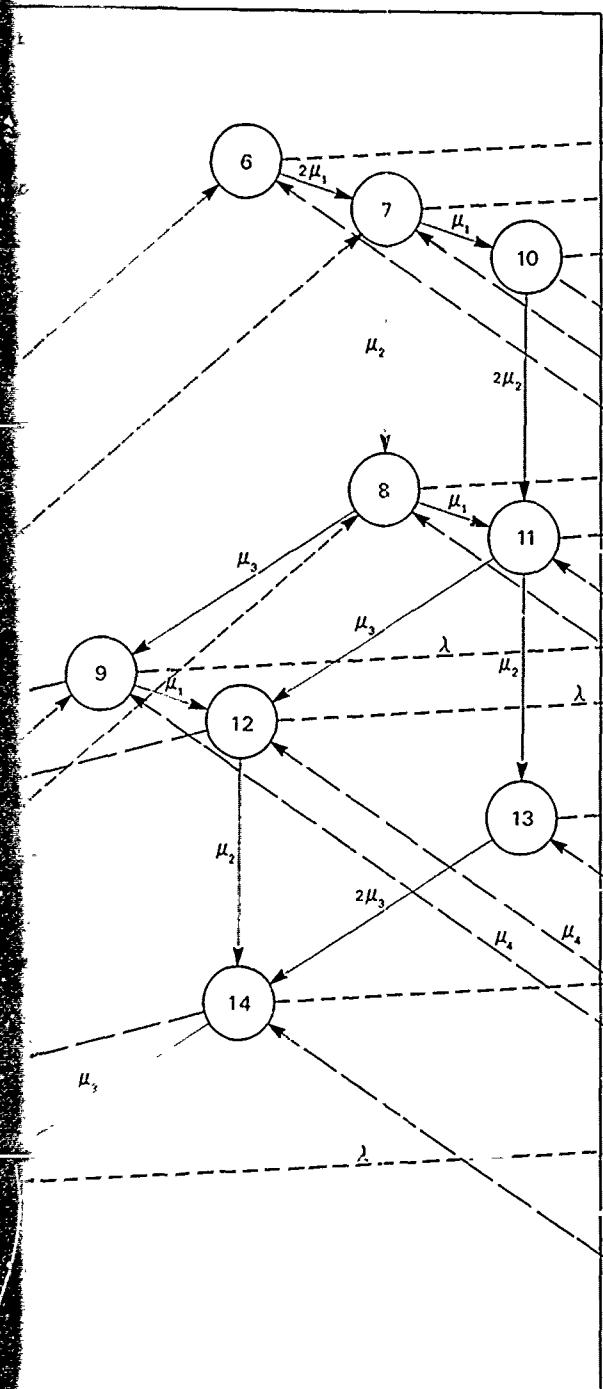


FIGURE III-25 CONFIGURATION OF FLOW GRAPH FOR UP TO THREE IN SYSTEM EXHIBITING PERIODIC

THREE IN SYSTEM

TWO IN SYSTEM



EXHIBITING PERIODICITY OF LOOP TRANSITIONS HAVING A PERIOD OF FIVE STEPS

The values of u_1 to u_4 used for the unsaturated system time distribution calcualtion are given in Figure III-24. The value of λ was taken to be 3.5. The calculated (cumulative) probability curve for unsaturated time is given in Figure III-26.

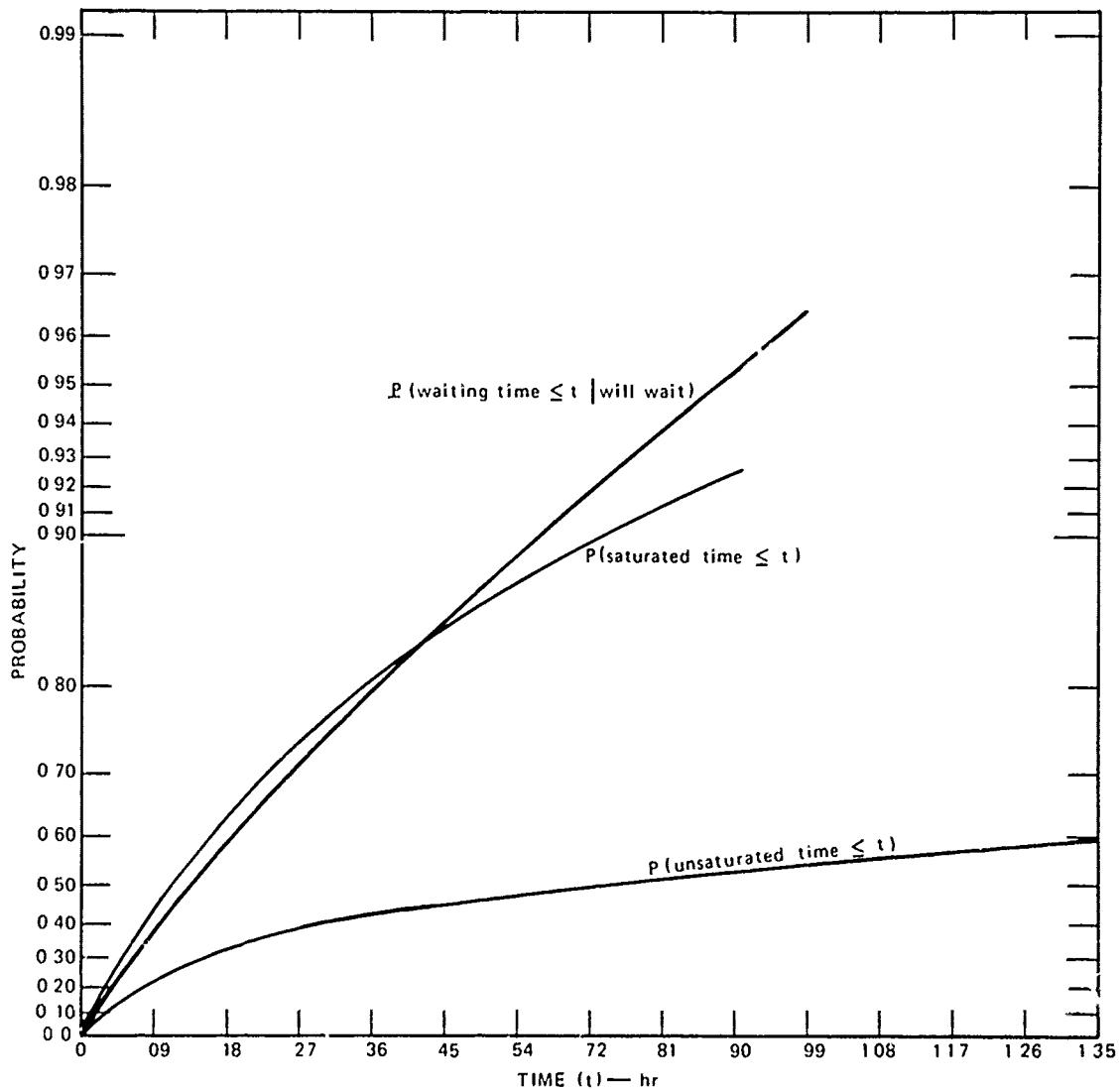


FIGURE III-26 PROBABILITY DISTRIBUTIONS FOR UNIT MODEL

Basically, the same approach was used for saturated and waiting time as for unsaturated time, except that much larger flow graphs had to be dealt with for the first two, forcing consideration of a maximum of

5 stimuli in the first queue, Q_D , whereas up to 11 stimuli were considered in Q_D for the unsaturated time case. The reason for this was a desire to adhere to a reasonable total CPU* time constraint per run, and run time is more or less proportional to total number of states of the model. As stated earlier, the method of integration to date is not yet refined for efficiency; rather, the main concern was relative ease of programming for the purposes of experimentation with the prototype model. Better integration methods are discussed in the next section. The results of computations for saturated time and waiting time are included in Figure III-26. Note that, if these distributions were actually exponential as assumed in Section IV, their curves would be straight lines. Thus, one can see from Figure III-26 that the simplifying assumptions of Section IV are not too unrealistic, especially for waiting time.

In ASW terms, then, Figure III-26 shows the probability of:

1. A stimulus having to "wait" a time up to "t" hours (t hours or less) before detection capability is applied to that stimulus
2. The system's being saturated (no detection availability) for up to t hours
3. The system's having continuous detection availability for up to t hours.

E. Areas for Further Research

1. Numerical Calculations

There are several alternative approaches to numerical calculation of both the steady state and transient queuing models employed thus far. Some alternatives have already been mentioned, such as folding the

* Central processor unit in a computer.

flow graph for steady state and using a direct instead of an iterative linear equation subroutine in numerical integration.

Some additional alternatives to be investigated are flow graph reduction through algebraic manipulation by computer and/or direct calculation of flow graph transmission through use of "gain formulas," such as those of Mason, Coates, and the more recently introduced Chan-Mai flow graph.¹¹

2. Extension of Model to Include Nonexponential Statistics

The model has so far been restricted to having exponential distribution for all transition probabilities between states of the system. As pointed out in Appendix A, this amounts to creating a Markov process. The general theory of flow graphs extends easily to semi-Markov processes, which would result from allowing more general interstate transition distributions. Numerical calculation techniques for semi-Markov flow graphs have yet to be investigated. In the case of rational Laplace transforms in a semi-Markov flow graph, it is possible to represent each arc in it by a subnetwork of arcs having a transmission of the type $a/(s+b)$, thereby creating a higher dimensional Markov process. For more general semi-Markov flow graphs rational approximations can be used. Alternative approaches to be examined may include available techniques for direct integration of the integro-differential equations that arise for a semi-Markov process. The Laplace transformed version of these is given in Equation (A-17) of Appendix A.

The flow graph methodology fortunately lends itself as a convenient general approach to formulation of integro-differential equations. The literature on the subject of integral equations pertaining to queuing theory consists of a great "bag of tricks," each individually derived for a very specific problem. Thus far, it has not been found possible to find

a direct application of papers on integral equations in queuing theory as stated in Section III-A, but it appears likely that the unifying flow graph technique will lead to satisfactory solution of the integral equations arising for more general models than so far considered.

3. Addition of Classification and Hand-off Functions

The above remarks on integro-differential equations and flow graphs are pertinent to extension of the unit model to include various other ASW functions omitted so far. These functions can be introduced by compounding the dimensionality of the type of Markov or semi-Markov model employed to date. This approach will probably involve consolidating states of the resultant highly dimensional semi-Markov process and approximating it by a simpler, lower dimensional process. There has been a great deal of work done in systems analysis on approximation of functions by simpler rational functions, and this area of techniques should be investigated.

IV ASW FORCE EFFECTIVENESS MODEL

The purpose of this section is to describe the development of the ASW force effectiveness model. In this case effectiveness is defined as the probability that an ASW stimulus will be detected by at least one of the force sensors in sufficient time for effective counteraction by units and/or operating units of the force. In this model the effective counteraction of the operating units is not explicitly represented, but is implied through the incorporation of certain statistical results from the unit model.

In addition to a description of the model formulation, illustrative computer results are presented and the areas and direction for future model expansion are indicated.

A. Model Development

Consider a naval force which includes a certain number of ASW and/or operating units. For the sake of simplicity, operating units will be referred to hereinafter as units except when ambiguity might result.

For the purposes of this model, consider that the ASW elements are described as a collection of $i (i=1, 2, 3, \dots, m)$ ASW sensors, some of which may be collocated on the same platform. In addition, there may be duplications of the various sensor types within the force since the i indexing serves only to enumerate the sensors and not to categorize them. Let (x_o, y_o) be the position of the force reference point, either force center or position and intended movement (PIM). Since all velocity vectors in the development that follows are considered to represent motion relative to this reference point, there is no loss of generality if (x_o, y_o) is

assumed to be positioned at the origin of a rectangular coordinate system (see Figure IV-1). Let the position of the i^{th} sensor relative to the force reference position within this coordinate system be (x_i, y_i) .

In a similar manner, let the position of the r^{th} stimulus ($r=1, 2, 3, \dots$) relative to the force reference position be $(x_s, y_s)_r$. In most of the discussion that follows, the r -index will be suppressed for the sake of notational simplicity, so that stimuli positions will be referred to merely as (x_s, y_s) . The identity of the specific stimulus under discussion should be clear from the context of the discussion, but, if this is not the case, the r -index notation will be employed. Other quantities are defined as indicated:

DEFINITION OF SYMBOLS

S	The set of all ASW-significant stimuli.
T	A subset of S consisting of all submarine related stimuli.
F	A subset of S consisting of all non-submarine related stimuli.
S_k ($k=1, 2, \dots, n^1$)	A sequence of subsets of S each of which consists of stimuli which exhibit identical ASW related characteristics. By construction:

$$S_j \cap S_k = \emptyset \quad j \neq k$$

$$\bigcup_k S_k = S .$$

A member of a particular subset, S_k , may be referred to as, "a k -type stimulus."

It should also be noted that for some k_1 , k_2 , and k , it may be that:

$$k_1 \in S_k \quad k_2 \in S_k$$

$$k_1 \in T \quad k_2 \in F .$$

$S(i)$	A subset of S consisting of all stimuli for which detection by the i^{th} sensor is possible.
--------	---

R_{ik}^* Maximum detection range of the i^{th} sensor for a k -type stimulus (see Figure IV-1).

\bar{R}

The radius of the force area of interest
(see Figure IV-1).

$g_{ik}(R)$

The single look probability that the i^{th} sensor will detect a k -type stimulus given that the stimulus is at range R and the detection capability of the i^{th} sensor is not saturated. Applies to sensors which employ a "glimpsing" type of operation.

$\gamma_{ik}(R)$

The instantaneous probability density of detection of the i^{th} sensor for a k -type stimulus given that the stimulus is at range R and the detection capability of the i^{th} sensor is not saturated. Applies to sensors which employ a "continuous looking" type of operation.

WR

Submarine weapon range (see Figure IV-1).

$v_i = (\theta_i, v_i)$

The velocity vector of the i^{th} sensor (operating element) relative to the force reference point. v_i is described by a heading (direction of relative movement), θ_i , measured counterclockwise from the positive x -axis, and a velocity, v_i .

$v_s^r = (\varphi_s, v_s)_j$

The velocity vector of the r^{th} stimulus relative to the force reference point.

v_s^r is described by a heading, φ_s , measured counterclockwise from the positive x -axis, and a velocity, v_s . As previously stated, the r -index will normally be suppressed.

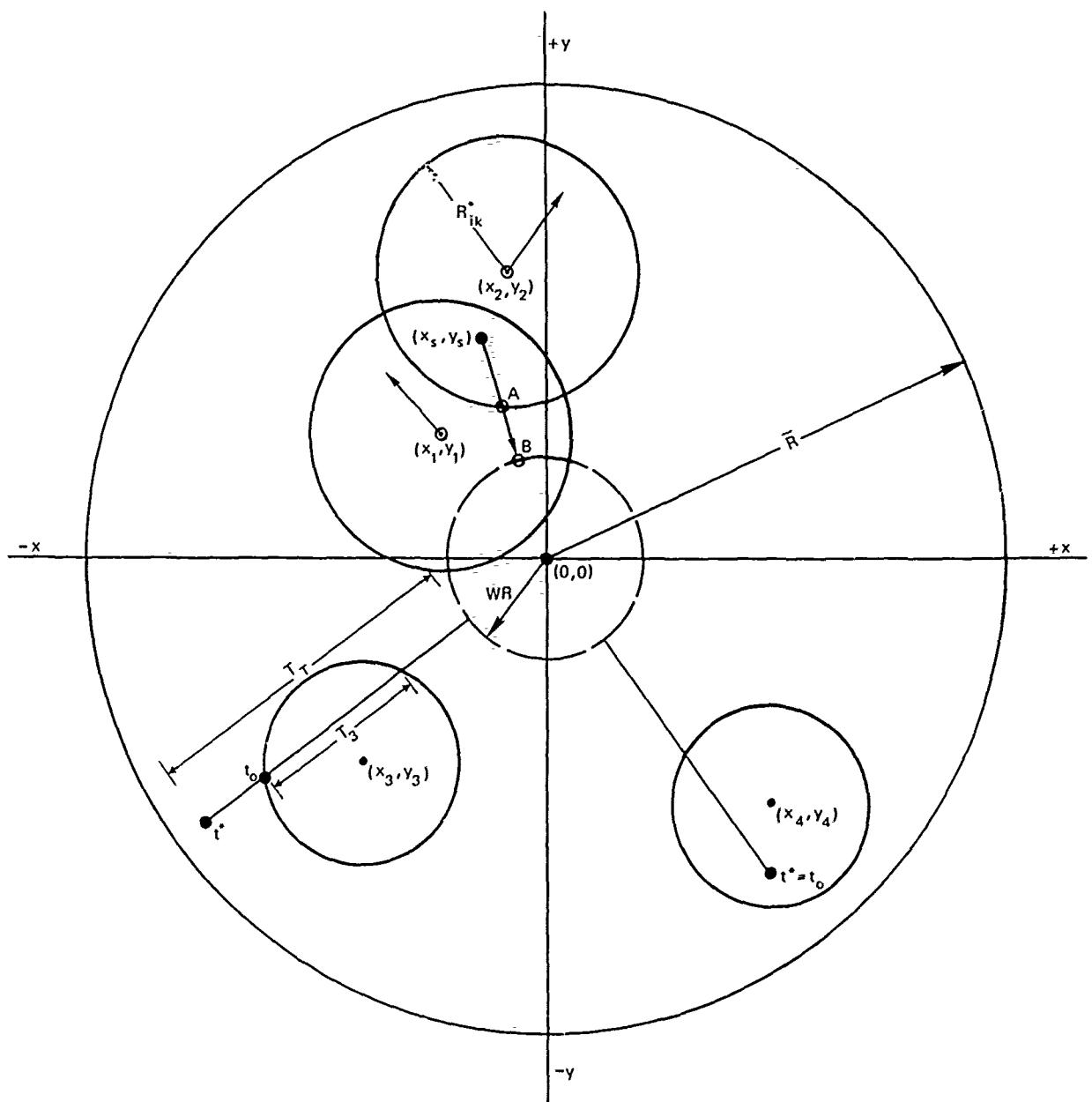


FIGURE IV-1 SIMPLIFIED DIAGRAM OF SENSOR AND STIMULUS RELATIONSHIPS

Now consider the force at time t^* (see Figure IV-1). M stimuli have been presented to the units of the force up to time t^* of which L stimuli have become contacts ($L \leq M$). Of these L contacts, ℓ are currently being prosecuted by the units of the force and $\ell(i)$ are being prosecuted by the unit of which the i^{th} sensor is a part. Obviously,

$$\ell = \sum_i \ell(i)$$

and

$$0 \leq \ell(i) \leq \ell \quad \text{for all } i$$

which constitutes the load on the force and individual systems respectively at time t^* .

At time t^* the $(M+1)^{st}$ stimulus is presented to the force, that is, the $(M+1)^{st}$ stimulus appears within range R of the force center. The questions to be addressed are: (1) will the $(M+1)^{st}$ stimulus become a contact within the ASW system? and (2) if so, will there be sufficient time for effective counteraction by the elements of the force? The degree to which these two questions collectively are answered in the affirmative constitutes the force level measure of effectiveness. Clearly, the first question may be interpreted in terms of the probability that at least one of the force sensors will detect the $(M+1)^{st}$ stimulus; the second question concerns the time required for detection.

With regard to this latter point, for a given stimulus initial position, velocity vector, and weapon range, there is a maximum time, T_T , within which a "true" stimulus must be countered. In Figure IV-1, the time T_T would correspond to the time required for the stimulus to transmit from (x_s, y_s) to the point B on the weapon range boundary. In this and the subsequent model development, it is assumed that the defended portion of the force does not take evasive action to avoid a contact. Such

action introduces the concepts of action and counteraction between the defended force and the contact, which are beyond the scope of the model at this time.

The model does, however, incorporate provisions for specifying the movement of the ASW units. Such movement is dictated by the ℓ contacts presently in the system and is not due to the presence of the $(M+1)$ st stimulus. This consequence is obvious and stems directly from the fact that only initial detection of the $(M+1)$ st stimulus is being considered. As a result, the force is unaware of the presence of this stimulus. It is also possible, under actual operational conditions, that the motion of the ASW units will be in response to the occurrence of contacts subsequent to the appearance of the $(M+1)$ st stimulus. This leads to more complicated provisions for describing the relative stimulus-unit movement, which are not yet incorporated within the present model structure.

At time t^* assume that the $(M+1)$ st stimulus is located at (x_s, y_s) and is proceeding in accordance with a velocity vector, v_s , relative to the force reference point. Assume, too, that the position and velocity vector, (x_i, y_i) and v_i , are known for each of the i sensors. Recall that it is possible that

$$(x_i, y_i) = (x_j, y_j)$$

and

$$v_i = v_j$$

for some i and j due to the enumeration feature of the i -index and the possibility of sensor collocation. In addition, let a submarine weapon range, WR , be established about the units of the defended force to

delineate the limit for possible ASW counteraction. This is the situation depicted in simplified form in Figure IV-1.

Consider the i^{th} sensor and assume that the $(M+1)^{\text{st}}$ stimulus is a k' -type stimulus such that

$$(M+1) \in S_{k'} \subset S(i) \quad (\text{IV-A-1})$$

and

$$R_{ik'} < R_{ik}^* \quad (\text{IV-A-2})$$

for at least some portion of the relative stimulus track within \bar{R} . These two conditions are assumed in order to insure that the i^{th} sensor is potentially capable of contributing positively to the effectiveness of the force. If either condition is violated, the contribution of the i^{th} sensor is obviously nil.

It is quite possible that the $(M+1)^{\text{st}}$ stimulus will not be within detection range of the i^{th} sensor at t^* . However, under the assumption contained in (IV-A-2), there exists an initial time t_0 (see Figure IV-1) and a subsequent period when the stimulus will be within detection range of the i^{th} sensor. This period, designated T_i ($i=1, 2, 3, \dots, m$), is the maximum detection interval available to the i^{th} sensor. The actual detection interval available to the i^{th} sensor will now be determined.

Assuming that the $(M+1)^{\text{st}}$ stimulus originates from an attacking submarine, the time at which the stimulus reaches the weapon range boundary is

$$t^* + T_T ,$$

and the amount of time available to the i^{th} sensor in which to accomplish detection and effective counteraction is

$$T_T(i) = T_T + t^* - t_o \quad (i=1,2,3,\dots,m) \quad (\text{IV-A-3})$$

where, by definition $t_o \geq t^*$.

Let $\tau_i (i+1,2,3,\dots,m)$ be the time required by the ASW unit of which the i^{th} sensor is a part to prosecute a contact. τ_i is a random variable which may be contact dependent and most certainly is dependent on relative contact-unit geometry. Assume that the distribution function, $F_{\tau_i}(z)$, such that

$$F_{\tau_i}(z) = P[\tau_i \leq z]$$

for the system load $\ell(i)$ is known, and that the statistical properties of this distribution are not significantly altered by the detection of the $(M+1)^{\text{st}}$ stimulus and its subsequent introduction into the system. From a practical viewpoint it is expected that $F_{\tau_i}(z)$ will be determined from the unit model and input to the force effectiveness model.

Two situations relating the i^{th} sensor and the $(M+1)^{\text{st}}$ stimulus are possible, depending upon the relationship that exists between the weapon range circle about the defended force and the detection field of the i^{th} sensor. These two conditions are illustrated in Figure IV-1. Referring to Figure IV-1, the first sensor, located at (x_1, y_1) , must detect and counter the stimulus during its transit from (x_s, y_s) to point B. On the other hand, the second sensor, located at (x_2, y_2) , need only detect the stimulus during its transit from (x_s, y_s) to point A, assuming, of course, that there is sufficient time for effective counteraction before the stimulus reaches point B.

Now the time for the stimulus to reach point A is $t^* - t_o + T_2$ and the time to reach point B is $t^* - t_o + T_1$. For the first sensor the actual detection interval available $(DI)_1$ is given by

$$(DI)_1 = t^* - t_o + T_1 - \tau_1 = T_T(1) - \tau_1$$

through the use of (IV-A-3). The corresponding detection interval for the second sensor, $(DI)_2$ is

$$(DI)_2 = t^* - t_o + T_2$$

assuming that $\tau_2 \leq (t^* - t_o + T_2) - (t^* - t_o + T_1) = T_2 - T_1$.

Generalizing these two conditions, the actual detection interval available for the i^{th} sensor is

$$(DI)_i = \begin{cases} T_i & t^* + T_T - \tau_i \geq t_o + T_i \\ t^* - t_o + T_T - \tau_i & t^* + T_T - \tau_i < t_o + T_i \end{cases}$$

or, more concisely,

$$(DI)_i = \min[T_i, T_T(i) - \tau_i] \quad (i=1, 2, 3, \dots, m) \quad (\text{IV-A-4})$$

where $T_T(i)$ is defined by (IV-A-3), T_i is the total time the stimulus remains within detection range of the i^{th} sensor, and it is assumed that $T_T(i) - \tau_i > 0$.

Before proceeding, a discussion of some of the relevant features of the unit model is necessary. In its present form, the unit model is a queuing model of the ASW functions of a unit consisting of multiple

service stages in series with multiple service facilities in most of the stages, such as, the detection stage. However, the unit model considers only the time necessary to accomplish detection, not the probability that detection will occur. In order to develop some of the statistical quantities which describe the functioning of a unit, a hypothetical queue is assumed to exist ahead of the detection function. Clearly, this does not represent actual conditions since stimuli, particularly submarine stimuli, do not "wait" on the ASW system. The use of this artifice does permit, however, the determination of such quantities as: (1) the probability that a stimulus that encounters the i^{th} sensor at time t will not have to wait, which is equivalent to the probability that the detection server is available; and (2) The probability that the stimulus will wait for a time less than some time, say t' , given that it waits at all. This latter probability is equivalent to the probability that the remaining portion of the detection server saturated period is less than t' , given that the server is saturated. In this context, a saturated period is defined to be the period throughout which the detector server is fully occupied. By contrast, an unsaturated period is defined to be the period when some portion of the system detection capability is available to acquire new targets. Because of the probabilistic nature of detection, it is quite possible that a detector server may be unsaturated even though appropriate stimuli are present and within range.

Let X_w , X_s , and X_u be random variables representing, respectively, the length of a stimulus-customer's wait in the detection queue, the length of a detection server saturated period, and the length of a detection server unsaturated period. Assume that the distribution functions $F_{X_w}(t)$, $F_{X_s}(t)$, and $F_{X_u}(t)$ are known as a result of exercising the unit model. Then, by definition of a distribution function:

$$P[X_w \leq t] = F_{X_w}(t) = \int_0^t f_{X_w}(x)dx$$

$$P[X_s \leq t] = F_{X_s}(t) = \int_0^t f_{X_s}(x)dx$$

$$P[X_u \leq t] = F_{X_u}(t) = \int_0^t f_{X_u}(x)dx$$

As previously stated, $F_{X_w}(t)$ for $t > 0$, represents the probability that the remaining portion of the current detector server saturated period will last at most t time units after the arrival of a stimulus-customer. Obviously, a stimulus may arrive during a detector unsaturated period and so there is a positive probability that the initial waiting period will be of length zero. This fact is utilized in the force model by considering that the sequence of saturated and unsaturated periods of detector server activity faced by a stimulus always starts with a saturated period, albeit a period possibly of length zero. Subsequent saturated periods, so far as a particular stimulus is concerned, do not exhibit this characteristic.

Implicit in this last discussion is the assumption that the detector server is of finite capacity. It is apparent that a server with an exceedingly large or infinite capacity might never experience a saturated period, but, under present practical operating restraints, such servers are not likely to be encountered.

The procedure for determining the probability that the i^{th} sensor will detect the $(M+1)^{st}$ stimulus during the detection interval $(DI)_i$ is as follows.

Define a sequence of time points, $\{t_k\}$ ($k=1, 2, 3, \dots, n$), such that $t_0 < t_1 < t_2 \dots < t_n$. From this sequence of time points, construct a sequence of intervals, $\{I_k\}$ ($k=1, 2, 3, \dots, n$), with length, $L(I_k)$, equal to $t_k - t_{k-1}$. It is assumed that the $L(I_k)$ are independent random variables with the distribution function:

$$P[L(I_j) \leq t] = \begin{cases} F_{X_w}(t) & j=1 \\ F_{X_s}(t) & j=2k-1, k=2, 3, \dots, n/2 \\ F_{X_u}(t) & j=2k, k=1, 2, \dots, n/2 \end{cases} .$$

Since only sequences, $\{I_k\}$, which terminate with an unsaturated period are of practical importance, the number, n , of intervals in the sequence will always be even.

Now, during the $(DI)_i$ the $(M+1)$ st stimulus will be exposed to the i^{th} sensor during periods when the i^{th} sensor is unsaturated and when it is saturated. Of course, detection of the $(M+1)$ st stimulus can take place only during one of the unsaturated periods. This condition raises questions regarding the number, spacing, and length of the i^{th} sensor unsaturated periods.

Define the two sums, S_j^s and S_j^u as follows

$$S_j^s = \sum_{k=1}^{2j-1} L(I_k) \quad j = 1, 2, 3, \dots, n/2 \quad (\text{IV-A-5})$$

$$S_j^u = S_j^s + X_{u_j} \quad j = 1, 2, 3, \dots, n/2 \quad (\text{IV-A-6})$$

where $X_{u_j} \equiv L(I_{2j})$. The quantity, S_j^s , represents the sum of the lengths of the first j saturated periods and the first $j-1$ unsaturated periods.

Whereas, s_j^u represents the sum of the first j saturated and unsaturated periods alike.

The density functions of the two sums in (IV-A-5) and (IV-A-6) may be derived by applying a well-known rule of probability theory regarding random sums. This rule states that the density function of a random sum of independent random variables is the convolution of the density functions of all of the random variables comprising the sum. Thus,

$$f_{s_1^s}(x) = f_{I_1}(x) = f_{X_w}(x)$$

$$f_{s_1^u}(x) = f_{s_1^s}(x) * f_{X_u}(x)$$

$$f_{s_2^s}(x) = f_{s_1^u}(x) * f_{X_s}(x)$$

$$f_{s_2^u}(x) = f_{s_2^s}(x) * f_{X_u}(x)$$

or, in general

$$f_{s_1^s}(x) = f_{X_w}(x) \quad (IV-A-7)$$

$$f_{s_j^s}(x) \approx f_{s_{j-1}^u}(x) * f_{X_s}(x) \quad j=2, 3, \dots, n/2 \quad (IV-A-8)$$

$$f_{s_j^u}(x) = f_{s_j^s}(x) * f_{X_u}(x) \quad j=1, 2, 3, \dots, n/2 \quad (IV-A-9)$$

where (*) denotes convolution.

Let d_{ij} be the probability that the i^{th} sensor detects the $(M+1)^{\text{st}}$ stimulus in the j^{th} unsaturated period and not before. This probability is determined as follows.

The j^{th} unsaturated period commences at some time, t' , where for $d_{ij} > 0$, t' , in absolute time, must be in the interval $t_o \leq t' \leq t_o + (DI)_i$. Since only the i^{th} sensor is of concern in this discussion, there is no loss of generality if t' is considered relative to the i^{th} sensor detection period so that the range of values for t' is $0 \leq t' \leq (DI)_i$. The length of the j^{th} unsaturated period is equivalent to $L(I_{2j})$ and $0 \leq L(I_{2j}) \leq (DI)_i - t'$. In the interest of future notational simplicity, let:

$$L(I_{2j}) \equiv y . \quad (\text{IV-A-10})$$

Let $P_D(y|t')$ be the probability that the i^{th} sensor will detect the $(M+1)^{\text{st}}$ stimulus in an interval of length y given that the interval starts at t' . Assume for the present that $P_D(y|0)$ is given by

$$P_D(y|0) = \int_0^y f_D(x)dx$$

so that

$$P_D(y|t') = \int_{t'}^{t'+y} f_D(x)dx . \quad (\text{IV-A-11})$$

Then, $d_{ij}(y, t')$, the probability that the $(M+1)^{\text{st}}$ stimulus is detected by the i^{th} sensor in the j^{th} unsaturated period commencing at t' and of length y and not before, is

$$d_{ij}(y, t') = P_D(y|t')$$

. $P[j^{\text{th}}$ unsaturated period is of length $y|t']$

. $P[j^{\text{th}}$ unsaturated period starts at $t']$

. $P[\text{no detection in first } j-1 \text{ unsaturated periods}]$

and

$$\begin{aligned} d_{ij} &= \int_0^{(DI)_i} dt' \int_0^{(DI)_i - t'} d_{ij}(y, t') dy \\ &= \int_0^{(DI)_i} dt' \int_0^{(DI)_i - t'} P_D(y|t') \cdot f_{X_u}(y) \cdot f_{S_j}(t') \cdot \bar{P}_D(j-1) dy \\ &= \left[\int_0^{(DI)_i} dt' f_{S_j}(t') \int_0^{(DI)_i - t'} P_D(y|t') \cdot f_{X_u}(y) dy \right] \bar{P}_D(j-1) \end{aligned}$$

(IV-A-12)

where $\bar{P}_D(j-1)$ is the probability of no detections in the first $j-1$ unsaturated periods.

The probability, D_{ij} , that the $(M+1)^{\text{st}}$ stimulus is detected by the i^{th} sensor at some time during the first j unsaturated periods is

$$D_{ij} = \sum_{l=1}^j d_{il} \quad j=1, 2, 3, \dots, n/2 \quad (\text{IV-A-13})$$

with

$$\bar{P}_D(0) \equiv 1$$

$$\bar{P}_D(j-1) = 1 - D_{i,j-1} = 1 - \sum_{l=1}^{j-1} d_{il} \quad j=2,3,\dots,n/2 . \quad (\text{IV-A-14})$$

In particular, the probability that the i^{th} sensor detects the $(M+1)^{\text{st}}$ stimulus is:

$$D_i = \sum_{j=1}^{n/2} d_{ij} .$$

It should be noted that, due to the inherent physical restraints of the engagement situation, D_i is likely to be less than unity with the probability of the i^{th} sensor failing to detect being equal to $1 - D_i = \bar{P}_D(n/2)$.

Recall from (IV-A-4) that $(DI)_i$ is the minimum of either T_i or $T_T(i) - \tau_i$. If $T_i < T_T(i) - \tau_i$ then the expression in (IV-A-12) gives the desired probability since T_i , and so $(DI)_i$, for the stated conditions is a known or determinable value. If, however, $0 < T_T(i) - \tau_i < T_i$, the length of the detection interval, $(DI)_i$, becomes a random variable and the d_{ij} of (IV-A-12) should be written as $d_{ij}(DI)$, i.e., as a function of the length of the detection interval. In this case, d_{ij} would be determined by:

$$d_{ij} = \left[\int_0^{T_T(i)} f_{DI}(z) \cdot d_{ij}(z) dz \right] \bar{P}_D(j-1) . \quad (\text{IV-A-15})$$

It is necessary, in this case, to derive the probability density function, $f_{DI}(z)$, for the length of the detection interval. Therefore, for:

$$(DI)_i = T_T(i) - \tau_i \quad (\tau_i \text{ a random variable})$$

$$\begin{aligned} F_{(DI)_i}(z) &= P[(DI)_i \leq z] = P[T_T(i) - \tau_i \leq z] = P[\tau_i \geq T_T(i) - z] \\ &= 1 - F_{\tau_i}(T_T(i) - z) = 1 - \int_0^{T_T(i)-z} f_{\tau_i}(x) dx . \end{aligned}$$

Differentiating,

$$f_{(DI)_i}(z) = f_{\tau_i}(T_T(i) - z)$$

so that (IV-A-15) becomes

$$d_{ij} = \left[\int_0^{T_T(i)} f_{\tau_i}(T_T(i) - z) d_{ij}(z) dz \right] \bar{P}_D(j-1) \quad (\text{IV-A-16})$$

since $f_{\tau_i}(z)$ is assumed known from the unit model, and $d_{ij}(z)$ represents the value of the expression in the brackets on the right-hand side of (IV-A-12).

In the present development, no consideration is given to either maneuvering to increase T_T [and thus $T_T(i)$] or to hand-off's between operating units. Obviously, in real situations it is possible for $T_T(i) - \tau_i < 0 < T_i$. This, quite naturally, gives rise to questions regarding the existence of another unit with prosecution time of, say,

τ_j , where $(x_i, y_i) \neq (x_j, y_j)$, i.e., the j^{th} sensor is not a part of the same unit as the i^{th} sensor and such that:

$$T_T(i) - \tau_i < 0 < T_T(i) - \tau_j .$$

This also suggests that, from the force point of view, it might be possible to find a unit containing the j^{th} sensor such that, for the $(M+1)^{\text{st}}$ stimulus:

$$\tau_j = \min_i (\tau_i) .$$

As stated, these and other associated questions have not yet been addressed.

Continuing with the model development, the probability that at least one of the i sensors comprising the ASW force will detect the $(M+1)^{\text{st}}$ stimulus in time for effective counteraction is

$$D = 1 - \prod_{i=1}^m (1 - D_i) = 1 - \prod_{i=1}^m \left(1 - \sum_{j=1}^{n/2} d_{ij}\right) \quad (\text{IV-A-17})$$

where, of course, many of the D_i may be identically zero since either the $(M+1)^{\text{st}}$ stimulus is not of the proper type: $(M+1) \in S_k \subset S(i)$ or $R_{ik} > R_{ik}^*$ for all of the relative stimulus track within \bar{R} . The quantity D constitutes the ASW force measure of effectiveness.

The probability D in (IV-A-17) should be written as $D(x_s, y_s, v_s)$ since the quantity D is actually a function of the initial position of the $(M+1)^{\text{st}}$ stimulus and its assumed velocity vector. If the position (x_s, y_s) were allowed to vary over all possible feasible positions within \bar{R} (with v_s assumed constant for all stimuli), the value D would trace out a surface which would characterize the detection capability of the force. A

series of these surfaces for various stimulus velocity vectors, v_s , could serve to diagnose the ASW detection weaknesses of the selected force.

It remains to develop the functional form of the probability that the i^{th} sensor will detect the $(M+1)^{st}$ stimulus in an interval of length y , i.e., the $P_D(y|t')$ in (IV-A-11). Two cases must be considered: the continuous looking type of sensor and the intermittent or "glimpsing" type of sensor. In both cases it is assumed that a probability of detection versus range curve appropriate to the sensor type, its operational style and the pertinent environmental conditions exists and is known. This curve, functionally, is either $g_{ik}(R)$ or $\gamma_{ik}(R)$ depending upon the sensor style of operation. It should be noted that, since sensor and stimulus initial positions and velocity vectors are assumed known, both $g_{ik}(R)$ and $\gamma_{ik}(R)$ are, in fact, continuous functions of time since the range between the sensors and the stimulus may be expressed as a continuous function of time.

Consider the continuous looking case. Koopman¹² has shown that the probability of detection in this case is given by

$$P_D(y|t') = 1 - e^{-F[c]} \quad (\text{IV-A-18})$$

where

$$F[c] = \int_{t'}^{t'+y} \gamma_{ik}[R(t)] dt \quad (\text{IV-A-19})$$

and the integral in (IV-A-19) is a line integral along the path traced by the $(M+1)^{st}$ stimulus relative to the i^{th} sensor. The expression, $\gamma_{ik}[R(t)]$, is employed in (IV-A-19) to show explicitly the functional dependence of γ upon time.

Obviously, (IV-A-18) and (IV-A-19) result in a continuous function in the two variables t' and y . For a given system with known function $\gamma_{ik}(R)$, (IV-A-18) and (IV-A-19) derive the function $P_D(y|t')$, which appears in (IV-A-11) and (IV-A-12).

In the glimpsing case, the probability of detection during the j^{th} unsaturated period is a function of the number of glimpses during this period. Accordingly, the quantity $d_{ij}(y,t)$ in (IV-A-12) is replaced by the quantity $d_{ij}(x)$ where x is the number of glimpses in the j^{th} unsaturated period. $d_{ij}(x)$ is determined as follows.

Let $t_L > 0$ represent the time for a single glimpse. Then,

$$\begin{aligned} d_{ij}(x) &= P[\text{detection in } x \text{ glimpses} | t'] \\ &\cdot P[x \text{ glimpses during } j^{\text{th}} \text{ unsaturated period} | t'] \\ &\cdot P[j^{\text{th}} \text{ unsaturated period starts at } t'] \cdot \bar{P}_D(j-1) \end{aligned}$$

where the last two probabilities are identical to those appearing in the definition of $d_{ij}(y,t')$.

The probability that there are x glimpses during the j^{th} unsaturated period is equivalent to the probability that the length of the j^{th} unsaturated period is such that

$$\left[xt_L \leq y < (x+1)t_L \right]$$

given that the j^{th} unsaturated period commences at some arbitrary point t' . In terms of the density function, $f_{X_u}(y)$, this probability is:

$$= \begin{cases} 0, & \text{if } t' + xt_L > (DI)_i \\ \int_{xt_L}^{(x+1)t_L} f_{X_u}(y) dy, & \text{if } t' + xt_L \leq (DI)_i \end{cases} . \quad (\text{IV-A-20})$$

The probability that the detection occurs in x glimpses, given that the unsaturated period starts at t' , is given by

$$1 - \prod_{n=1}^x \left[1 - g_{ik}(t' + nt_L) \right] \quad (\text{IV-A-21})$$

where the dependence of the functional value of $g_{ik}(\cdot)$ upon the starting point of the unsaturated period is made explicit by the argument. Clearly, the function in (IV-A-21) is a function of two variables: the number of glimpses and the starting point of the unsaturated period t' . For any specified value of x , the function in (IV-A-21) is a continuous function in the variable t' over the interval $(0, (DI)_i - xt_L)$. This fact will be exploited in the formulation of $d_{ij}(x)$.

Combining the foregoing,

$$d_{ij}(x) = \left[\int_0^{(DI)_i - xt_L} dt' f_{S_j^s}(t') \left[1 - \prod_{n=1}^x (1 - g_{ik}(t' + nt_L)) \right] \left[\int_{xt_L}^{(x+1)t_L} f_{X_u}(y) dy \right] \right] \bar{P}_D^{(j-1)} \quad (\text{IV-A-22})$$

and

$$d_{ij} = \sum_{x=1}^{\lceil (DI)_i / t_L \rceil} d_{ij}(x) \quad (\text{IV-A-23})$$

where $\lceil (DI)_i / t_L \rceil$ in (IV-A-23) signifies the largest integer that is less than or at most equal to the quantity in the brackets.

If, as discussed previously, $(DI)_i = \tau_i$, a known quantity, then (IV-A-22) and (IV-A-23) give the desired probability. If, however, $(DI)_i = T_T(i) - \tau_i$, then (IV-A-22) and (IV-A-23) must be modified as follows.

The $d_{ij}(x)$ in (IV-A-22) becomes $d_{ij}(x, DI)$ --the subscript on the DI having been suppressed--and $d_{ij}(x)$ is then given by

$$d_{ij}(x) = \left[\int_0^{T_T(i)} f_{\tau_i}(T_T(i)-z) d_{ij}(x, z) dz \right] \bar{P}_D(j-1) \quad (\text{IV-A-24})$$

where $d_{ij}(x, z)$ represents the value of the expression within the outermost brackets on the right hand side of (IV-A-22) and:

$$d_{ij} = \sum_{x=1}^{\lceil T_T(i) / t_L \rceil} d_{ij}(x) \quad (\text{IV-A-25})$$

The remainder of the development proceeds as previously described.

B. Illustrative Results

To illustrate the application of the force model a computer program was developed and the effectiveness of a representative ASW force determined.

The representative ASW force is depicted in Figure IV-2. The force consists of a defended portion (unspecified in nature), located at (0,0) and proceeding in a direction parallel to the positive y-axis, which is accompanied by eight ASW units. The location and direction of movement of the ASW units are shown in Figure IV-2 and detailed in Table IV-1. It should be recalled that both unit and stimulus velocities are relative to the velocity of the defended force. Thus, for element 1, a velocity of 0.0 knots indicates that this element is proceeding at the same pace as the defended force. The unit velocity vectors shown in Figure IV-2 represent only direction and not velocity. As can be seen from Figure IV-2

Table IV-1
ASW UNIT POSITIONS AND VELOCITY VECTORS

Unit Number	Position		Velocity Vector	
	x_i (nmi)	y_i (nmi)	θ_i (deg)	v_i (knots)
1	-20.0	+70.0	90	0.0
2	+20.0	+70.0	70	2.0
3	-60.0	+30.0	100	1.0
4	+60.0	+30.0	90	0.0
5	-60.0	-30.0	250	5.0
6	+120.0	-70.0	150	5.0
7	-40.0	-60.0	210	5.0
8	+70.0	-120.0	120	5.0

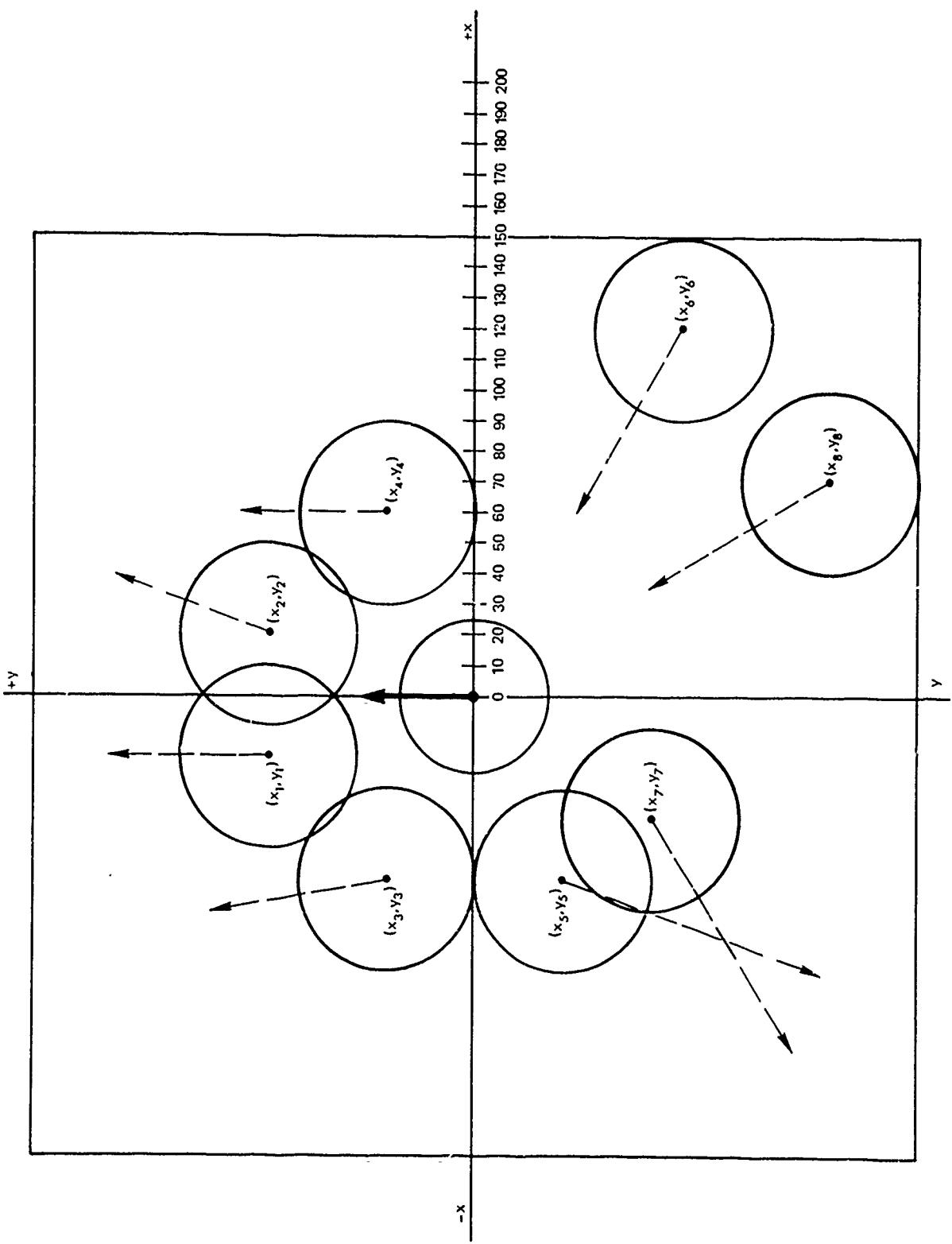


FIGURE IV-2 REPRESENTATIVE ASW FORCE FOR MODEL DEMONSTRATION

two units are rejoining the force (Nos. 6 and 8); two units are moving toward a contact (Nos. 5 and 7); and the remaining four are arranged in an advanced ASW screen. To illustrate the versatility of the model and program, only two of the units are moving in the same direction and at the same speed (Nos. 1 and 4).

For ease of presentation, the force area of interest, (\bar{R}), is assumed to be rectangular in nature and extending for 150 nmi either side of each major coordinate axis. The submarine weapon range consists of a circular area about the force center with a radius of 25 nmi which is roughly representative of possible cruise missile range. The stimuli are considered to originate (initial appearance) at each of the grid points contained within the force area of interest. In addition, each stimuli proceeds radially on an interception course at a speed of 10 knots--all relative to the defended force.

Each unit is considered to contain a single ASW sensor of the continuous looking type, hence expression (IV-A-12) from Subsection A above is applicable. The capability of each sensor is considered to be equivalent and is represented by a $\gamma_{ik}(R)$ of the following form

$$\gamma_{ik}(R) = a - bR^2$$

which is zero at $R = (a/b)^{1/2}$ and equal to a at $R = 0$. For this demonstration, the parameter values, a and b, are assumed equal to 0.9900 and 0.0011, respectively, so that the maximum detection range of the sensors (R_{ik}^*) is 30 nmi.

Employing this functional form of $\gamma_{ik}(R)$, the function $P_D(y|t')$ becomes

$$P_D(y|t') = \begin{cases} 0 & R > R_{ik}^* \\ 1 - e^{-F[c]} & R \leq R_{ik}^* \end{cases}$$

where

$$\begin{aligned} F[c] &= \int_{t'}^{t'+y} \gamma_{ik}[R(t)] dt \\ &= \int_{t'}^{t'+y} a - b \left\{ \left[x_s(t^*) - x_i(t^*) + t(\cos \varphi_s v_s - \cos \theta_i v_i) \right]^2 \right. \\ &\quad \left. + \left[y_s(t^*) - y_i(t^*) + t(\sin \varphi_s v_s - \sin \theta_i v_i) \right]^2 \right\} dt \end{aligned}$$

and the $x_i(t^*)$ and $y_i(t^*)$ of the ASW units are assumed to be the initial unit positions indicated in Figure IV-2.

The various probability density functions which characterize system performance are assumed to be of the following form

$$f_{X_W}(x) = P_o \delta(x-t_o) + (1-P_o) W e^{-Wx}$$

$$f_{X_u}(x) = I e^{-Ix}$$

$$f_{X_s}(x) = B e^{-Bx}$$

where $\delta(x-t_o)$ is the dirac delta function. The various parameter values that specify these functions are: $W = 0.15/\text{min}$, $B = 0.03/\text{min}$, $I = 0.01/\text{min}$, and $P_o = 0.64$ (P_o is the probability that the system is not saturated at time t_o).

The density function $f_{S_j^s}(x)$ is found by first determining the Laplace transforms of the convoluted functions which comprise this function, expanding this Laplace transform by means of partial fractions, and then determining the inverse Laplace transform of this expansion. This results in

$$f_{S_1^s}(x) = P_o \delta(x-t_o) + (1-P_o) W e^{-Wx}$$

$$f_{S_j^s}(x) = \sum_{m=1}^{j-1} \left[A e^{-Ix} \frac{x^{m-1}}{(m-1)!} + B e^{-Bx} \frac{x^{m-1}}{(m-1)!} \right]$$

$$+ (1-P_o) W \left[\frac{IB}{(I-W)(B-W)} \right]^{j-1} e^{-Wx}$$

where

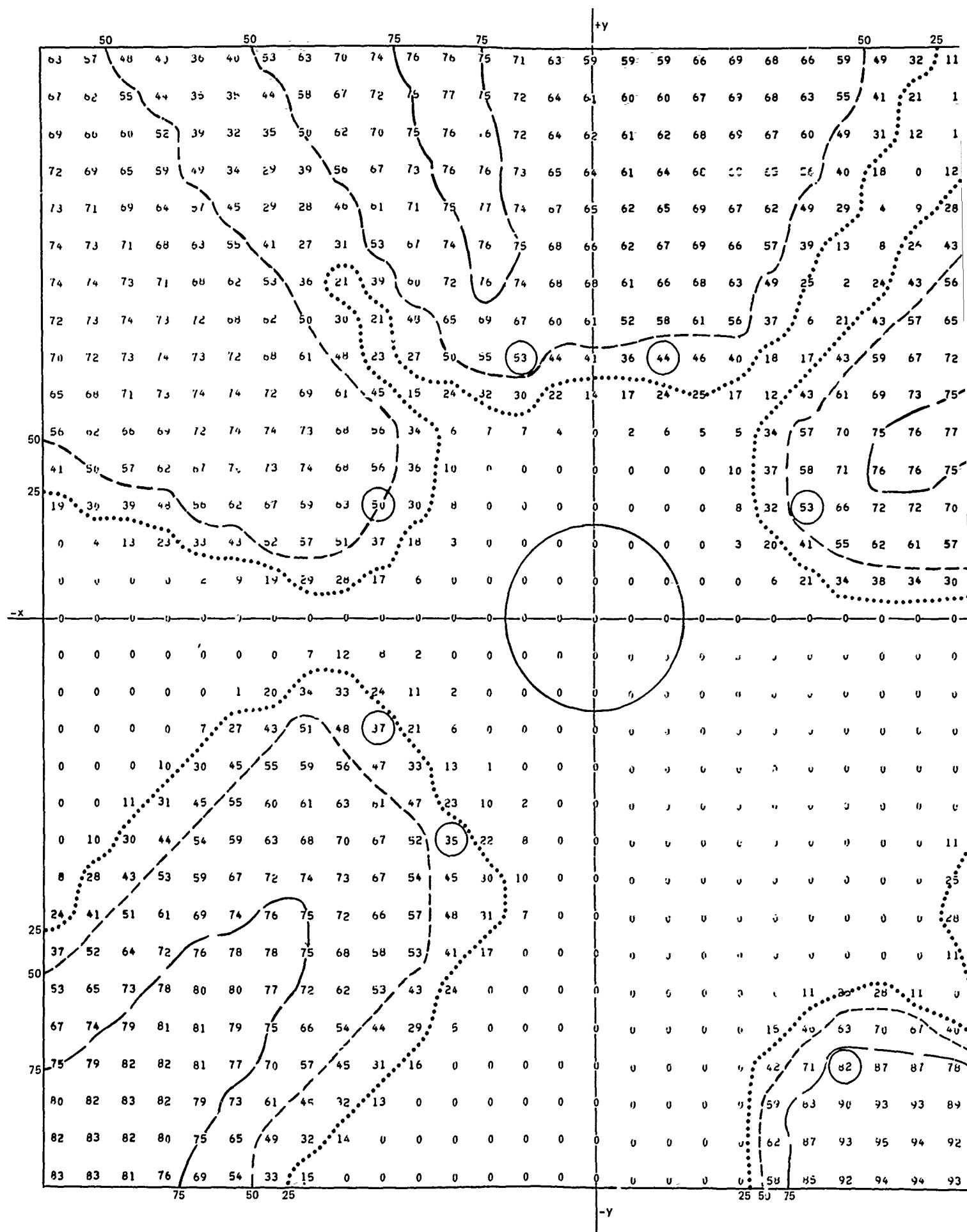
$$A = \frac{P_o [IB]^{j-1} (2j-m-3)! (-1)^{(j+m-1)}}{(B-I)^{(2j-2-m)} (j-2)! (j-m-1)!}$$

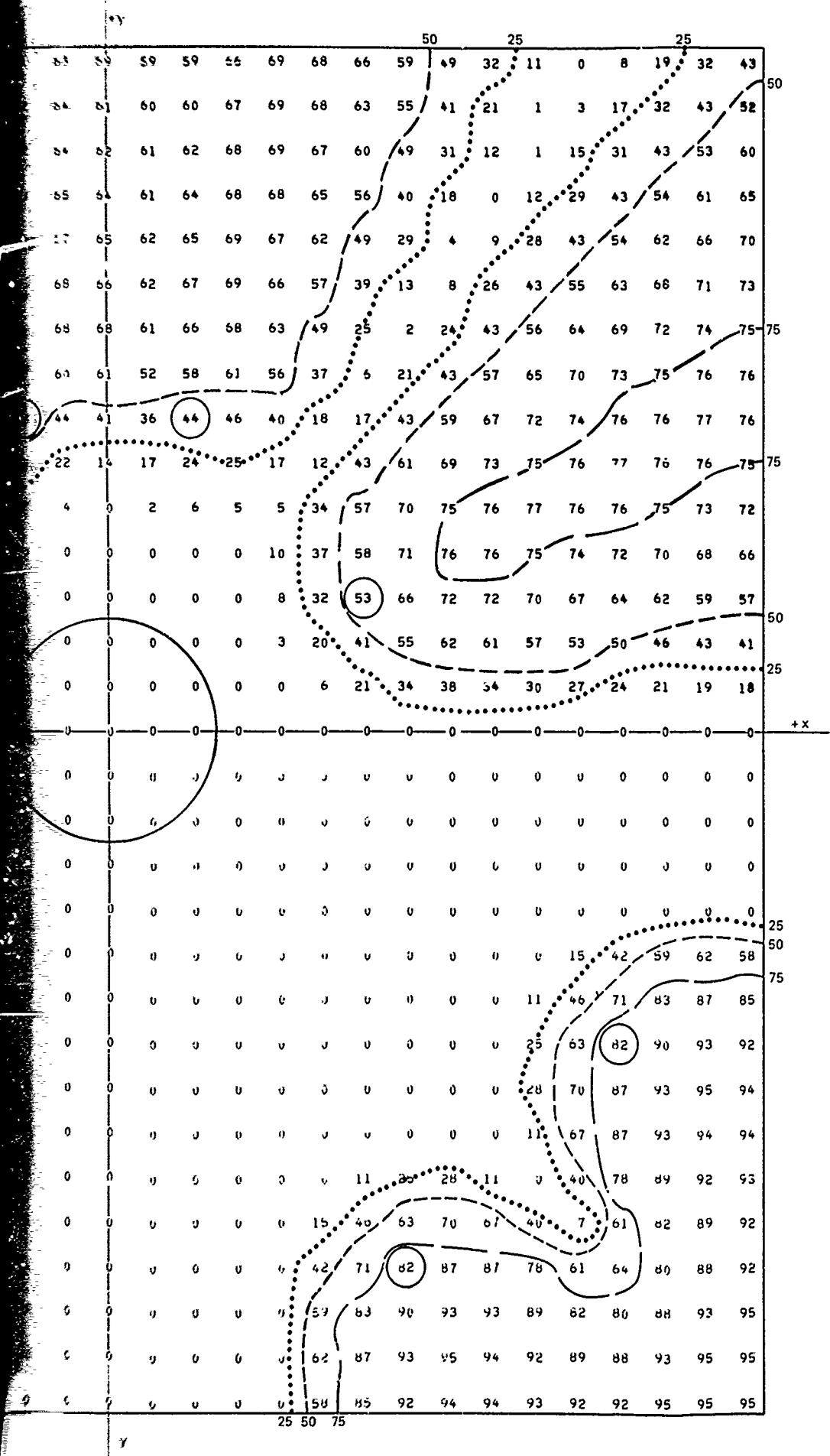
$$+ (-1)^{(j+m-1)} W (1-P_o) [IB]^{j-1} \sum_{i=1}^{j-m} \frac{(j+i-3)!}{(j-2)! (i-1)! (W-I)^{(j-m-i+1)} (B-I)^{(j+i-2)}}$$

$$B = \frac{P_o [IB]^{j-1} (2j-m-3)! (-1)^{(j+m-1)}}{(I-B)^{(2j-2-m)} (j-2)! (j-m-1)!}$$

$$+ (-1)^{(j+m-1)} W (1-P_o) [IB]^{j-1} \sum_{i=1}^{j-m} \frac{(j+i-3)!}{(j-2)! (i-1)! (W-B)^{(j-m-i+1)} (I-B)^{(j+i-2)}}$$

The results of the model calculations are presented in Figure IV-3 in the form of equal probability contours about the force center. In Figure IV-3 the initial positions of the ASW units and the weapon ranges about the defended force are indicated by the encircled numbers. In





**FIGURE IV-3 ILLUSTRATIVE FORCE
EFFECTIVENESS MODEL
RESULTS**

In addition, contours representing the probability that the $(M+1)$ st stimulus will be detected by at least one of the elements of the force are shown in the figure for the 0.25, 0.50, and 0.75 probability levels.

While the results shown in Figure IV-3 are for illustrative purposes only and do not represent the capability of any specific force, they do indicate the manner in which the ASW coverage is reduced if units either move out to investigate contacts (such as element 2) or are detached to prosecute contacts (such as elements 6 and 8). These results are, however, static in one sense. That is, although the kinematics among the units, the force, and the stimuli are explicitly included, the progression of time is not. Specifically, the detection capability afforded to the force while units 6 and 8 proceeded to their present position is not indicated. Thus, quite possibly, a large segment of the area in the lower right quadrant might have been "swept out" by these units while in the act of prosecuting their contacts. This implies that a time series of diagrams such as Figure IV-3 would be required, or, equivalently, some continuous elapsed time representation, in order to adequately describe the complete ASW capability of the force. Incorporation of this feature must be the subject of future research efforts. It is possible that this can be accomplished by treating the various grid point probabilities as a three-dimensional Markov process (x , y , and probability) and investigating the characteristics of such a process.

C. Directions for Continued Methodological Development

It has been shown that the ASW force effectiveness model represents the dynamic spatial and temporal relationships of an ASW engagement. The current stage of analytical development of the force effectiveness model should be extended, however, in order to incorporate more of the dynamic aspects of the ASW engagement. The purpose of this subsection is to

briefly describe the next logical developmental steps necessary to increase the dynamic applicability of this model.

1. Variable ASW Unit Motion

The prescribed ASW unit velocity vectors should be permitted to vary throughout the period of interest. The point of origin and length of the looking period in the model are allowed to vary throughout the entire range of the permissible values. To make the model more realistic, the ASW units should be allowed to alter their velocity vectors during the stimulus detection interval, thereby, altering the probability of detection versus range relationship to be considered; this would necessitate including an analytical representation of a partitioning of the variable length sensor looking period.

2. Intermittent and/or Combined Stimulus Presence

The purpose of this development is to introduce into the present model the effects on possible target detection due to either intermittent stimulus presence or multiple target stimuli. During the development of the force effectiveness model, the type of sensor considered was not made explicit, and, accordingly, it was assumed that the stimulus of concern was present throughout the entire period of detection opportunity. Yet, when specific sensor types are considered, questions regarding stimulus presence and the manner of operation become important because of the implicit sensor stimulus relationship. For example, if the specific ASW sensor to be considered is a periscope detecting radar, this sensor can be effective only during periscope exposure. Hence, when evaluating the potential contribution of this type of sensor to the overall force effectiveness, it is necessary to consider whether or not the appropriate stimulus will be present, and, if so, when and for how long a period.

Insofar as the total force detection effectiveness is concerned, there is no requirement that detection be accomplished by any one specific sensor type. In fact, supporting detections, particularly by different sensor types, contribute greatly to overall force effectiveness. With regard to this last point, nonsub phenomena seldom exhibit multiple characteristics similar to those of an actual submarine target.

The implications of introducing these two contingencies, intermittent and combined stimulus presence, into the model are: (1) the present detection methodology must be modified so as to reflect the interaction between stimulus presence and the sensor unsaturated periods for at least those force sensors deemed to be most relevant to force effectiveness; and, (2) the overall force effectiveness model will likely have to be restructured in the form of a network that combines the effects and results of each of the individual sensor models.

3. Intermittent Sensor Operation

Intermittent sensor operation is the unit counterpart to the intermittent stimulus presence discussed above. There are ASW sensors, most notably the helicopter dipping sonar, which are operated in an intermittent manner. The conditions surrounding the effective operation of these sensor types are often complicated by the large sensor relocation distances which can occur between successive search periods. The incorporation of the effects of intermittent sensor operation will require a modification to the present detection methodology similar to that described in IV-C-2 above.

4. Contact Hand-off Process

Contact hand-off process should be incorporated in the ASW force model. Two levels of this hand-off process must be considered. These are:

- The hand-off or transfer of contacts between units of the ASW force. An example of this is the transfer of surface escort active sonar contact to the helicopter such as the light airborne multipurpose systems (LAMPS) that conduct the final localization, attack and kill functions.
- The hand-off of contact to the ASW force from surveillance systems external to the ASW force. An example of this is the transfer of the fixed site acoustic surveillance system's (SOSUS) contact to the ASW force.

The first of the above has been discussed in Section III-E-3. The incorporation of the second hand-off process will come about during the formulation and implementation of the command and control function within the ASW Force queuing model. Hand-off process will influence the priority assignments given to contacts that are already in the queue and awaiting attention.

5. Other Areas for Analysis

There are other operational procedures which must ultimately be incorporated into the model if the methodology is to be complete. However, many of the circumstances surrounding these procedures have not as yet been investigated, and, therefore, the methodological implications are unclear. The areas of analysis associated with these operational procedures are listed here, without description, for the purpose of indicating the probable direction of future methodological development:

- Effect of loss of contact and regain contact procedures
- Contact conversion: procedures for transfer of contact from detection stimulus source to prosecution stimulus source
- Effect of variable stimulus motion (this problem may be encountered and examined when the problem of variable ASW unit motion is investigated).

V REFERENCES

1. Sacks, J., "Ergodicity of Queues in Series," *Annals of Mathematical Statistics*, Vol. 31, No. 3, 1960, pp. 579-588.
2. Saaty, T. L., Elements of Queueing Theory, p. 213, McGraw-Hill, New York, 1961.
3. _____, Elements of Queueing Theory, p. 198, McGraw-Hill, New York, 1961.
4. Lindley, D. V., "The Theory of Queues with a Single Server," *Proceedings of the Cambridge Philosophical Society*, Vol. 48, 1952, pp. 277-289.
5. Humphrey, T. L., "Applications of Multidimensional Markov Processes to Queueing Theory," Ph.D. Dissertation, Department of Electrical Engineering, Stanford University, 1966.
6. Howard, R. A., *Dynamic Probabilistic Systems*, Vols. I & II, John Wiley & Sons, Inc., New York, 1971.
7. Luenberger, D. G., "Spectral Decomposition of Linear Transformations in Finite-Dimensional Spaces," Institute in Engineering-Economic Systems Report, Stanford University, December 1963.
8. Chow, Y., and Cassignol, E., Linear Signal-Flow Graphs and Applications, John Wiley & Sons, Inc., New York, 1971.
9. Gue, R. L., "Signal Flow Graphs and Analog Computation in the Analysis of Finite Queues," *Oper. Res. Soc. of America Journal*, Vol. 14, No. 2, 1966.
10. Varga, R. S., Matrix Iterative Analysis, Prentice-Hall, Englewood Cliffs, New Jersey, 1962.
11. Ghosh, S. N. and Ghosh, P. K., "Flowgraphs and Linear Systems," *Int. J. Control*, Vol. 14, No. 5, pp. 961-973, 1971.
12. Koopman, B. O., Search and Screening, OEG Report No. 56, Operations Evaluation Group, Office of the Chief of Naval Operations, Navy Department, Washington, D.C., 1946, pp. 18-24.
13. Kristiansson, L., "On the Semi-Markov Process and its Block Diagram Representation With Applications to Aircraft Mission Analysis," Technical Note TN 55, SAAB Aircraft Company, Linköping, Sweden.

14. Kraut, S., and Hartung, J., "Elementary Semi-Markov and Markov Renewal Processes With an ASW Application," paper submitted to Operations Research Journal, August 28, 1967 (prepared at Grumman Aircraft Engineering Corporation).
15. Shamblin, J. E., "Analysis of Probabilistic Networks by Markov Chains," paper presented at the 39th National Meeting of the Operations Research Society of America, May 6, 1971 (prepared at School of Industrial Engineering and Management, Oklahoma State University).
16. Conolly, B. W., "A Probabilistic Theory of Anti-Submarine Warfare Models Developed in Terms of Congestion Theory," Technical Report No. 144, SACLANT ASW Research Centre, La Spezia, Italy, April 15, 1969.
17. Culbertson, D. D., "An Application of Semi-Markov Chains to ASW Tactical Systems," Research Contribution No. 118, Center for Naval Analyses, Arlington, Virginia, November 24, 1969.

Appendix A

FLOW GRAPHS AND SEMI-MARKOV PROCESSES
ASSOCIATED WITH A MULTIPLE PHASE QUEUING SYSTEM

Appendix A

FLOW GRAPHS AND SEMI-MARKOV PROCESSES ASSOCIATED WITH A MULTIPLE PHASE QUEUING SYSTEM

1. Transition Rate Diagrams

In the body of this report a particular type of directed graph, known as a transition rate diagram in Markov process theory, was introduced to represent states, and transitions between these states, in a multiple phase queuing system having exponential holding times and interarrival times. As pointed out there, more general holding and interarrival times can be considered by replacing nodes of the graph with appropriate subnetworks.

The concept of a flow graph is associative more generally with Markov and semi-Markov processes, of which this queuing situation is a special case. It is a common procedure in queuing analysis to "embed" a Markov process in the system being analyzed--that is, to visualize, as the heart of the system, a Markov process of the appropriate sort. Many of the results available on M/G/1 and GI/M/m queues are obtained using this approach. In the type of queuing model under study in this project, however, a lot more can be obtained using embedded Markov processes as exponential distributions are assumed throughout the network. In fact, the queuing model as formulated in Sections II-A and III-B, when viewed from the proper perspective, is nothing less than a continuous time Markov process. Also, certain results on semi-Markov processes are useful for interpretive purposes.

Markov and semi-Markov processes have been frequently employed to model stochastic aspects of dynamic systems,¹³⁻¹⁷ particularly in military systems analysis. The application of flow graphs results in the

formulation of a unifying approach to an otherwise very complex situation, as well as providing considerable generality and flexibility.

2. Markov and Semi-Markov Processes

In a Markov process, the transition times between states are all exponentially distributed and depend only on the state from which each transition originates. Basically, a Markov process is specified by:

- a. Transition probabilities $p_{ij} = \text{Prob}(\text{pass from state } i \text{ to state } j)$
- b. Holding rates λ_i , with $\text{Prob}(\text{stay in state } i \text{ at least time } t \text{ before transition}) = e^{-\lambda_i t}$.

A semi-Markov process, then, is specified by:

- a. Transition probabilities $p_{ij} = \text{Prob}(\text{pass from state } i \text{ to state } j)$
- b. Holding time probability densities $h_{ij}(t) = \text{Prob}(\text{time in state } i \text{ equals } t \text{ before passing to state } j | \text{will pass from state } i \text{ to stat. } j)$.

Thus, a Markov process is clearly a special case of a semi-Markov process.

Before it can be shown how the specific problem of this report is related to Markov and semi-Markov processes, one other concept used in semi-Markov and Markov theory must be discussed.

3. Competing Processes

Semi-Markov processes can be interpreted as processes in which a "race" is taking place between independent stochastic processes, each one governing the time that a transition will take from a state i to some state j , given the system is in state i at time zero. This is referred to as a competing process model.⁶ The assumption is made in this model that whenever "something happens," i.e., the system makes a transition, all the appropriate "competing processes" are reset like clocks and the first one that "rings an alarm" is the one that determines

the next transition. This "race" interpretation is based on defining:

$$g_{ik}(t) = \text{Prob}(\text{successor state to state } i \text{ is state } k, \text{ at time } t | \text{competition started at time 0}).$$

Then there exist integral formulas⁶ expressing $h_{ij}(t)$'s and p_{ij} 's in term of the given $g_{ik}(t)$'s. Similarly, given the $h_{ij}(t)$'s and p_{ij} 's, it is possible to define the $g_{ik}(t)$'s by implicit formulas, but in general these cannot be solved explicitly.

It turns out that a Markov process is the particular case of a semi-Markov process for which the competing processes are all exponentially distributed, i.e.,

$$g_{ik}(t) = \lambda_{ik} e^{-\lambda_{ik} t}, \quad t \geq 0. \quad (\text{A-1})$$

It is this fact that is of great value in the study of the multiple phase queuing system considered here. In fact, the servers and "arrival controller" of the ASW queuing system can be interpreted as "competing processes" which operate independently and govern the state transitions in the total network represented by Figures III-4 to III-8. The λ_{ik} 's are therefore all equal to one of the five quantities $\lambda, \mu_1, \mu_2, \mu_3, \mu_4$.

4. Multiple Phase Service Queuing in Terms of Competing Processes

Using the "competing process" interpretation, any multiple phase service queuing system, and in particular the single element model, can be regarded as a Markov process in the following way:

Take as the states of the Markov or semi-Markov process the states described in Section III-B. Then declare the arrivals and services to be competing processes. In the standard notation,⁶ a transition probability matrix P is thus obtained with entries

$$p_{ij} = \begin{cases} \lambda_{ij} / \sum_{\substack{k=1 \\ k \neq j}}^N \lambda_{ik}, & i \neq j \\ 0, & i = j \end{cases} \quad (A-2)$$

where N is the total number of states. In the ASW single element case, the λ_{ij} 's are always one of the quantities $\lambda, \mu_1, \mu_2, \mu_3, \mu_4$, depending upon what the "competing process" creates the transition from i to j , or else λ_{ij} is 0 if there is no such transition. Similarly, in the standard notation, the holding rates are

$$\lambda_i = \sum_{\substack{j=1 \\ j \neq i}}^N \lambda_{ij} \quad (A-3)$$

the holding rate (diagonal) matrix is

$$\Lambda = \text{diag } (\lambda_i) \quad (A-4)$$

and the transition rate matrix A with

$$a_{ij} = \begin{cases} \lambda_{ij}, & i \neq j \\ -\lambda_i, & i = j \end{cases} \quad (A-5)$$

is related to Λ and P by

$$A = \Lambda (P-I) \quad (A-6)$$

where I is the $N \times N$ identity matrix.

The model becomes a semi-Markov process when the arrivals or services are not exponential. However, by substituting suitable finite subnetworks of exponential servers for the state nodes to approximate the arrival or service holding behavior, a Markov process can again be produced, although at the expense of an enlarged number of states.

5. Flow Graphs for Markov and Semi-Markov Processes

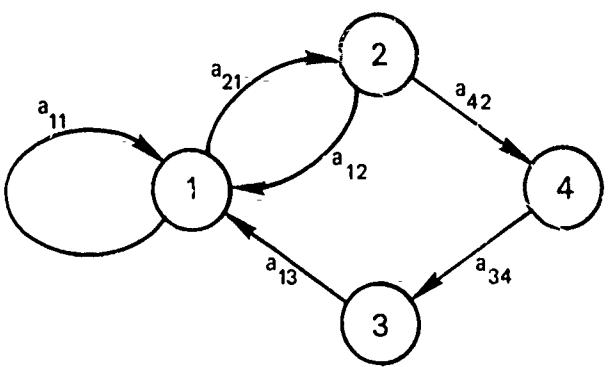
a. Flow Graphs

Flow graphs are particularly interpreted directed graphs associated with systems of linear equations.⁸ They were originally invented in signal flow theory of electrical engineering, and later worked their way into other areas, such as Markov and semi-Markov processes^{5,6,9} and queuing theory.

A flow graph for a system of linear equations

$$\sum_{j=1}^N a_{ij} x_j = x_i , \quad i=1, \dots, N \quad (A-7)$$

is interpreted as a network having nodes corresponding to the state variables x_i and having arcs labeled by a_{ij} directed from j to i for every nonzero a_{ij} (see Figure A-1). Note that the matrix $A-I$ has to be singular in order that the above equations can have a non-trivial solution. Thus, given a network with labels a_{ij} on its (directed) arcs, the system of equations it represents is always taken as the relations stating that each node's state variable x_i is obtained as the sum of adjacent state variables x_j at the tails of branches terminating at x_i multiplied by the arc weights a_{ij} of these branches, where j may possibly equal i . When $j=i$, this corresponds to a self-loop at node i (see Figure A-1).



$$\begin{aligned}
 x_1 &= a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\
 x_2 &= a_{21}x_1 \\
 x_3 &= a_{34}x_4 \\
 x_4 &= a_{42}x_2
 \end{aligned}$$

FIGURE A-1 TYPICAL FLOW GRAPH AND CORRESPONDING EQUATIONS

The linear equations represented by a flow graph can involve scalars, vectors, matrices, Laplace transforms, or matrices of Laplace transforms. As will be shown later, all four types will be useful in studying multiple phase queuing systems and Markov or semi-Markov processes in general.

Matrix flow graphs often provide a concise representation,⁶ especially when flow graphs are reducible to block form, in which it makes sense to distinguish between transitions within and between blocks of states.⁵

b. Transmission Through a Flow Graph

In flow graph theory there is the concept of "transmission" or "flow" through a flow graph having a pair of nodes singled out as "input" and "output" nodes. These are, of course, nodes that have only arcs leaving, or entering, respectively, with respect to the remainder of the flow graph (see Figure A-2). The motivation for the term "transmission" lies in signal flow graph theory; the formal definition is as follows:

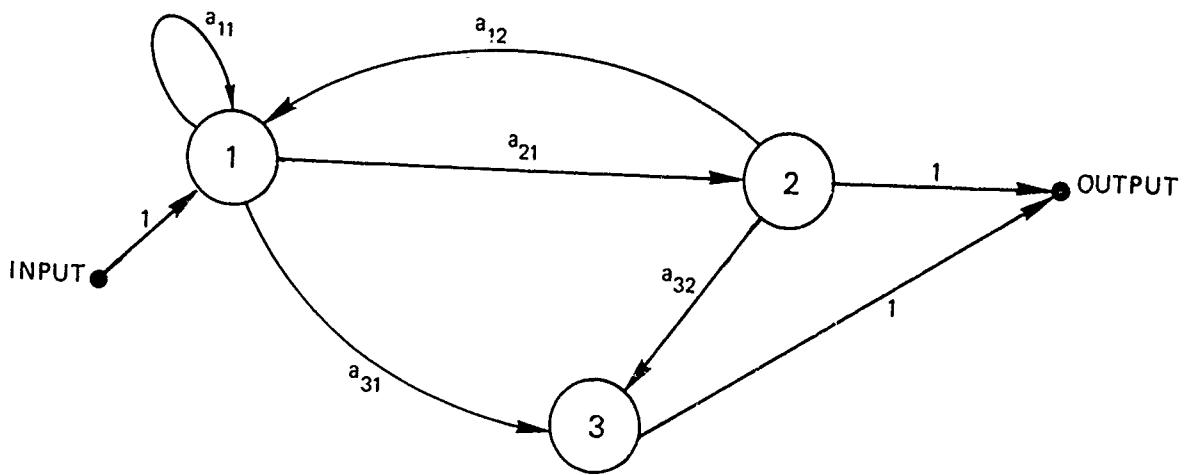


FIGURE A-2 TRANSMISSION THROUGH A FLOW GRAPH

Definition: Suppose that all variables except two particular ones in a flow graph have been algebraically eliminated, leaving one of these, referred to as the output variable, expressed as a factor times the other one, referred to as the input variable. Then this factor is referred to as the transmission of the flow graph from the designated input node to the designated output node.

c. Reduction and Equivalency of Flow Graphs

In the theory of flow graphs, there is an entire body of knowledge concerning reduction of flow graphs and equivalency of flow graphs. Two flow graphs are said to be equivalent, with respect to a set of input and output nodes, if they have the same transmission between these nodes. Nodes in a matrix flow graph are vectors, and thereby represent sets of nodes having scalars associated with them. The reduction techniques developed in signal flow theory were intended primarily for reducing flow graphs to a single arc from input to output so that the transmission could be calculated.

d. Steady-State Flow Graph of a Markov Process

The steady-state, or equilibrium, equations of a Markov process with transition probability matrix P and transition rate matrix Λ (see Section 4 of this appendix) are obtained from the dynamic equations (see Section e below) by setting derivatives equal to zero. This results in the matrix equation $p\Lambda = 0$, i.e.,

$$\lambda_i p_i = \sum_{j=1}^N p_{ji} p_j \lambda_j \quad i=1, 2, \dots, N \quad (\text{A-8})$$

where $p = (p_1, p_2, \dots, p_N)$ is a row vector of steady state probabilities

$$p_i = \text{Prob(system is in state } i \text{ in steady state)} \quad (\text{A-9})$$

assuming a steady-state exists. The flow graph for the above equations is given by Figure A-3. Note that the general a_{ij} of Figure A-2 contains the factor p_{ji} here, and so the p_{ij} arc is directed from i to j , and not from j to i .

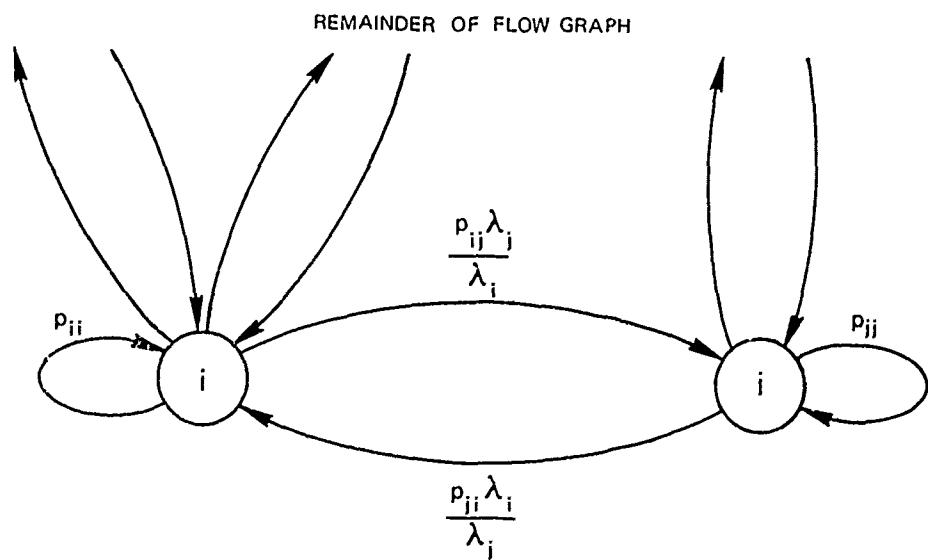


FIGURE A-3 STEADY-STATE FLOW GRAPH OF A MARKOV PROCESS

e. Dynamic Flow Graph of a Markov Process

In order to obtain a flow graph for a Markov process in the transient state, we must have a system of linear equations for this case, which is convenient to write in terms of Laplace transforms, and is given by

$$\varphi_{ij}^*(s) = \frac{\delta_{ij}}{s + \lambda_i} + \sum_{k=1}^N \frac{p_{ik} \lambda_k}{s + \lambda_k} \varphi_{kj}^*(s) \quad (A-10)$$

where $\varphi_{ij}^*(s)$ is the Laplace transform of the function $\varphi_{ij}(t)$ given by

$$\varphi_{ij}(t) = \text{Prob(system in state } j \text{ at time } t | \text{in state } i \text{ at time 0}) , \quad (A-11)$$

δ_{ij} is the Kroenecker delta function, and the p_{ik} 's and λ_k 's are as introduced in Sections 2 and 3 of this appendix.

In matrix form, the above equation may be written as

$$s\Phi^*(s) = \Phi^*(s)A \quad (A-12)$$

where A is defined by (A-6). Thus we have the matrix flow graph of Figure A-4 whose transmission, from an input vector of initial states to an output vector of terminal states, is a matrix of Laplace transforms given by

$$\begin{aligned} T &= (sI - A)^{-1} \\ &= \left[I - \frac{1}{s} A \right]^{-1} \cdot \frac{1}{s} I . \end{aligned} \quad (A-13)$$

The splitting of T into the above two factors explains the matrix self-loop in Figure A-4. The reason for resorting to this form is that it leads to a method of employing the coefficients of the matrix A directly. We cannot delve in depth into the general theory of flow graphs here, but will merely point out that self-loops with transmission G are commonly used to represent a "through" transmission equal to $(I-G)^{-1}$ (see Refs. 5,6).

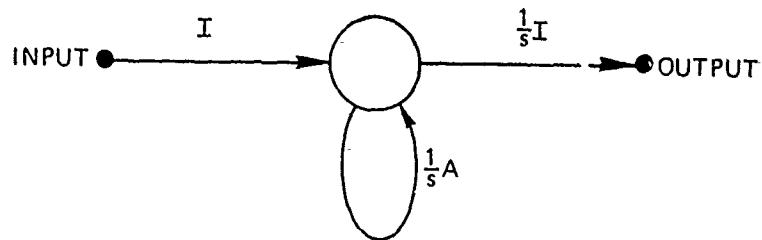


FIGURE A-4 MATRIX LAPLACE TRANSFORM FLOW GRAPH FOR A MARKOV PROCESS

The flow graph in Figure A-4 is useful from the theoretical standpoint, but for computations it must be broken down to a representation by components, as in Figure A-5. Note that in Figure A-5 there is a scalar input, i.e., the flow graph is entered externally at a single node, in order that the individual transmissions $\phi_{ij}^*(s)$ to the respective nodes may be computed, which yield the desired probability density functions

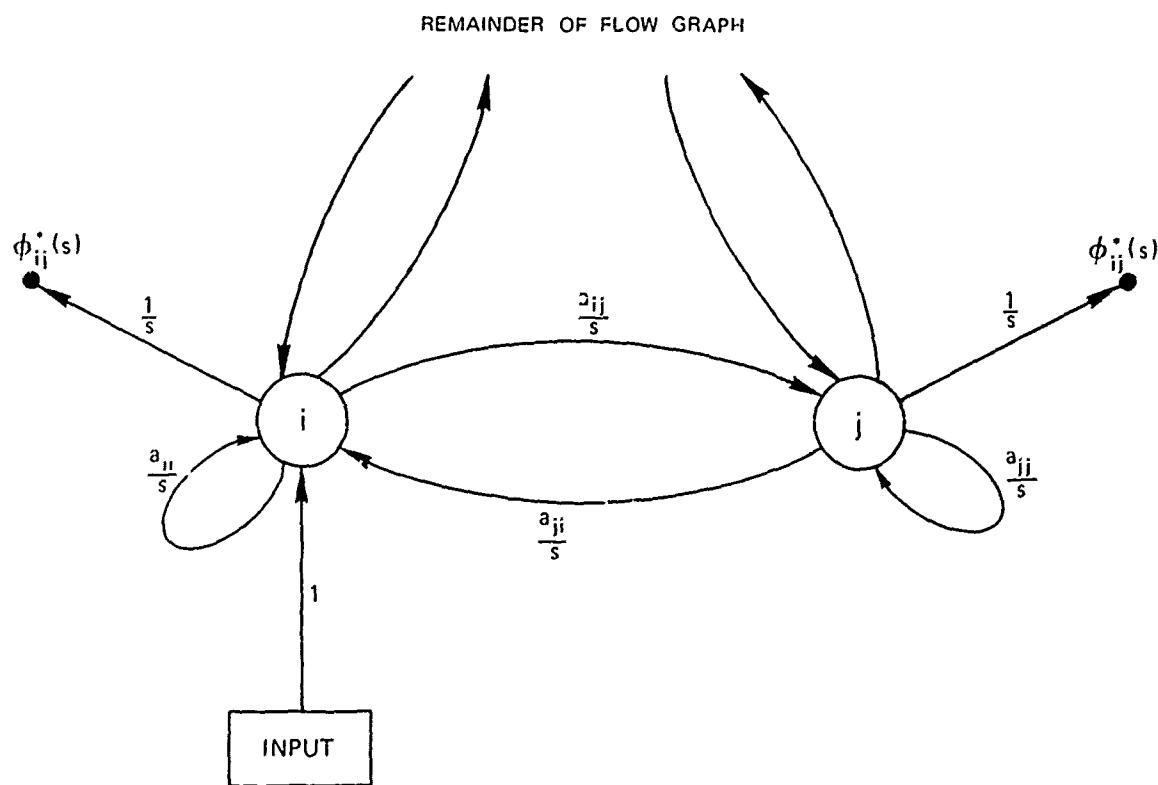


FIGURE A-5 SCALAR LAPLACE TRANSFORM FLOW GRAPH FOR A MARKOV PROCESS

$\varphi_{ij}(t)$ upon inversion of their transforms. Taking the input to the flow graph as node i corresponds to setting the initial conditions for $\varphi_{ij}(0) = \delta_{ij}$ in the corresponding time domain equation

$$\dot{\varphi}(t) = \varphi(t)A \quad (A-14)$$

of the Markov process; that is, $\varphi_{ij}(0) = \delta_{ij}$ is the condition that initially the system is in state i , which is required in the definition of $\varphi_{ij}(t)$.

In Figure A-5 the state variables at the nodes i are the quantities

$$x_j = s\varphi_{ij}^* \quad (A-15)$$

which are Laplace transforms of the derivatives of the desired probability distributions. The flow graph for the Markov process has been drawn with these state variables by Howard,⁶ and with the φ_{ij}^* 's themselves by Gue.⁹ Using

$$x_j = \varphi_{ij}^* \quad (A-16)$$

the only modification to Figure A-5 would be a factor of $1/s$ on the input arrow and a unity factor on the output arrows. The main point, however, is that the transmission from the input to each output is always the Laplace transform of the desired probability density function, regardless what equivalent network lies between. For computation by flow graph reduction techniques, the manner of choosing state variables may be important from the computational standpoint. Howard⁶ makes a great point of this in his treatment of flow graphs for particular probability distributions and he emphasizes the use of what he refers to as the semi-Markov flow graph for a Markov process, which is discussed below.

f. Dynamic Flow Graph of a Semi-Markov Process

The dynamic equations of a semi-Markov process, with state variables and parameters as defined in Section 5e of this appendix, are given, in terms of their Laplace transforms, by

$$\varphi_{ij}^*(s) = \delta_{ij} >_{w_i^*(s)} + \sum_{k=1}^N p_{ik} h_{ik}^*(s) \varphi_{kj}^*(s) \quad (A-17)$$

where φ_{ij} has the same meaning as for a Markov process, and:

$$>_{w_i}(t) = \text{Prob(amount of time system stays in state } i \text{ exceeds } t | \text{system in } i \text{ at time 0})$$

$$= \int_t^\infty \left(\sum_{j=1}^N h_{ij}(\tau) p_{ij} \right) d\tau \quad (A-18)$$

$$= e^{-\lambda_i t} \text{ in the Markov case .}$$

Howard's flow graph⁶ for the above equation appears in Figure A-6 in matrix form, and in Figure A-7 in scalar form.

It will be observed from comparison of Figures A-4 with A-6 and A-5 with A-7 that the state variables do not correspond in these two cases. In Figure A-7, the state variables are actually

$$x_j = \frac{1}{>_{w_j^*(s)}} \varphi_{ij}^*(s) \quad (A-19)$$

which reduces to

$$x_j = (s + \lambda_j) \varphi_{ij}^*(s) \quad (A-20)$$

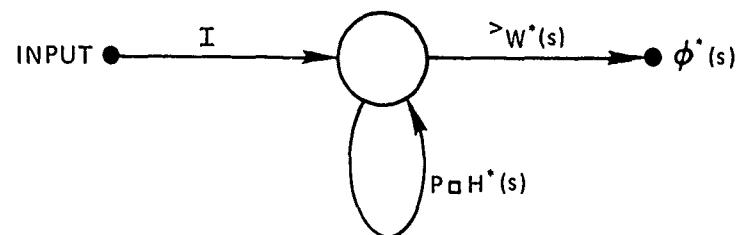


FIGURE A-6 MATRIX LAPLACE TRANSFORM FLOW GRAPH FOR SEMI-MARKOV PROCESS

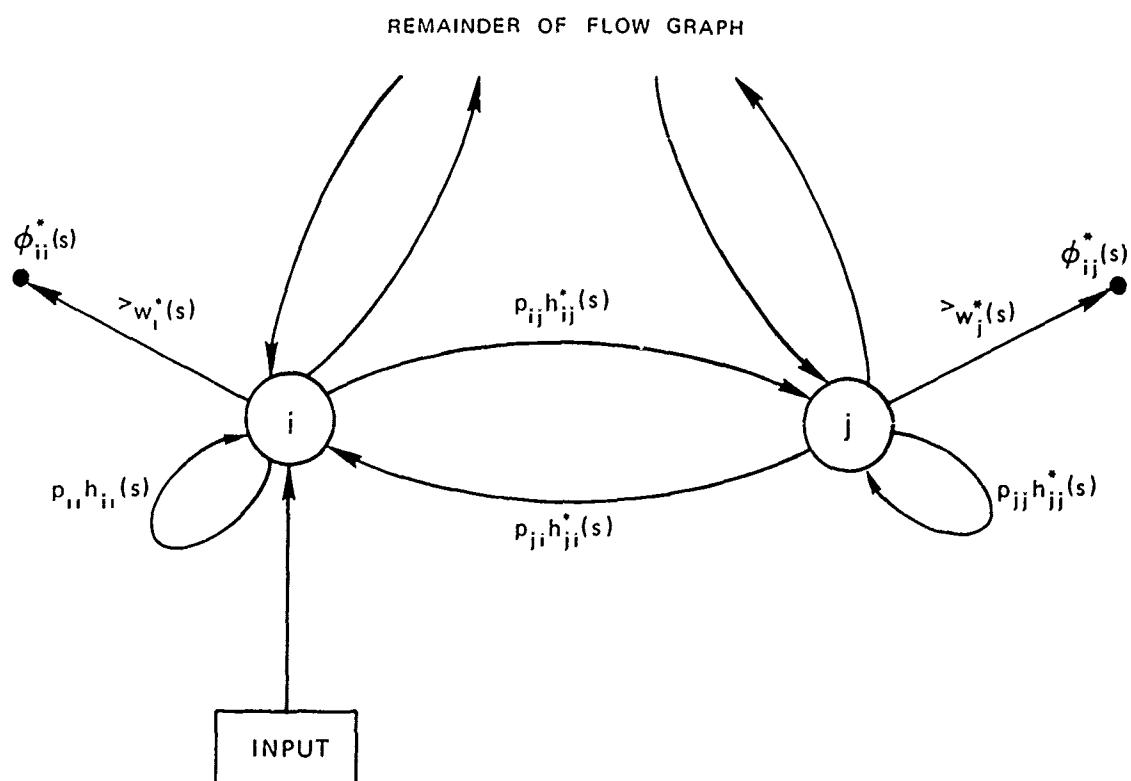


FIGURE A-7 SCALAR LAPLACE TRANSFORM FLOW GRAPH FOR A SEMI-MARKOV PROCESS

in the Markov case while, as was seen earlier, the Markov flow diagram has:

$$x_j = s \varphi_{ij}^*(s) . \quad (A-21)$$

The reason for this discrepancy is that Markov flow graphs have been conveniently based on the A matrix because it totally describes the process. Since no such simple representation exists for a semi-Markov process, the semi-Markov flow graphs have been based on the P and H matrices. As Howard points out,⁶ the semi-Markov flow graph of a Markov process has the advantage that it has no self-loops when the p_{ii} 's are all 0, which is often the case (as it is in this application). Thus, hand computations by flow graph reduction techniques are aided by resorting to this approach. The special case of Figure A-7 takes the form of Figure A-8 when we have a Markov process.

When a Markov process is expressed in terms of competing processes as in Section 3, the arc factors in Figure A-8 take the form:

$$p_{ij} \frac{\lambda_i}{s + \lambda_i} = \frac{\lambda_{ij}}{s + \sum_{j=1}^N \lambda_{ij}} \quad (A-22)$$

Thus, in the single element model, these will be of the form:

$$\frac{\mu_2}{s + \lambda + \mu_1 + \mu_2 + \mu_3 + \mu_4} , \text{ etc.} \quad (A-23)$$

The value of semi-Markov type flow graphs is not just to be able to eliminate self-loops in certain types of Markov flow graphs. Setting up semi-Markov flow graphs for queuing systems can be of direct

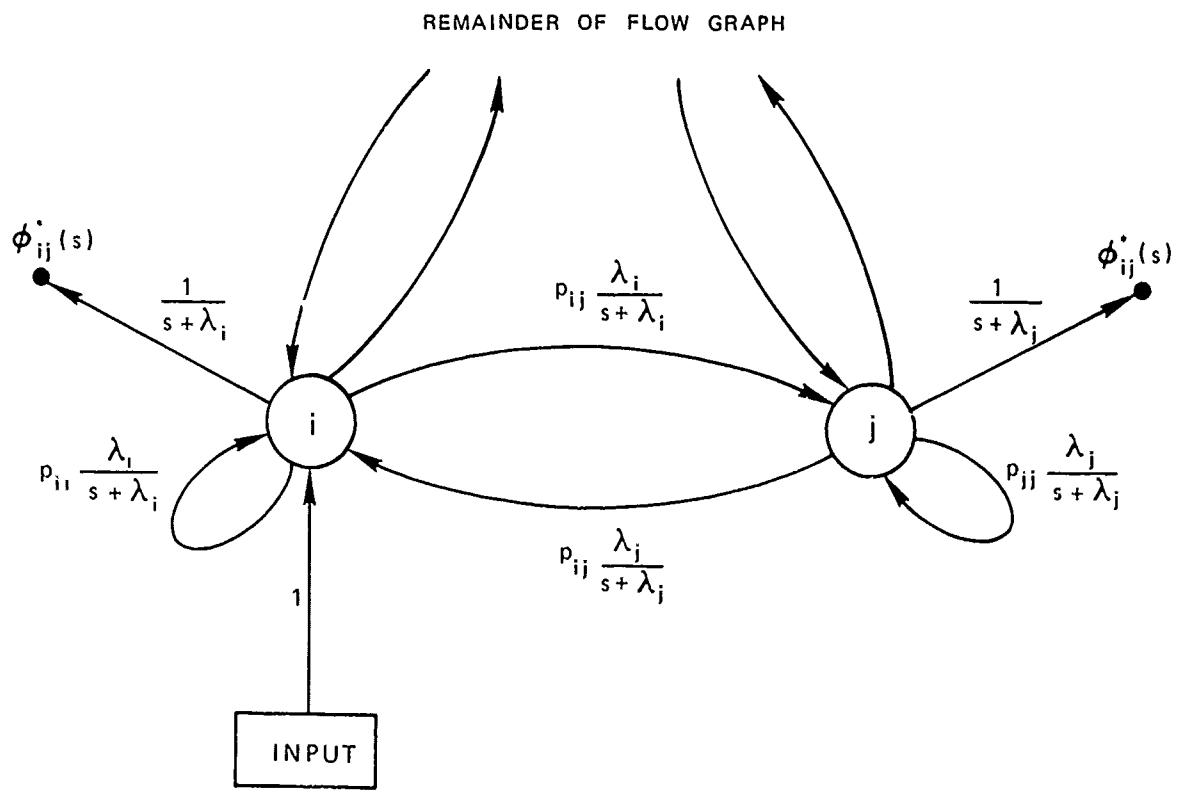


FIGURE A-8 SEMI-MARKOV FLOW GRAPH FOR A MARKOV PROCESS

value in numerical computation, when services and arrivals are not exponential (see Section III-E).

6. Using Semi-Markov Flow Graphs to Compute Probability Distributions for State Occupancy Times

Suppose it is desired to know the probability distribution for the amount of time spcnt between entering and exiting a particular subset S of the set of states in a Markov or semi-Markov process. This problem can be approached from the point of view of flow graph theory. Consider the subgraph of the semi-Markov flow graph for the given process consisting of only the nodes in S and all arcs interconnecting these nodes. Recall that the flow factors on the arcs here are Laplace transforms of

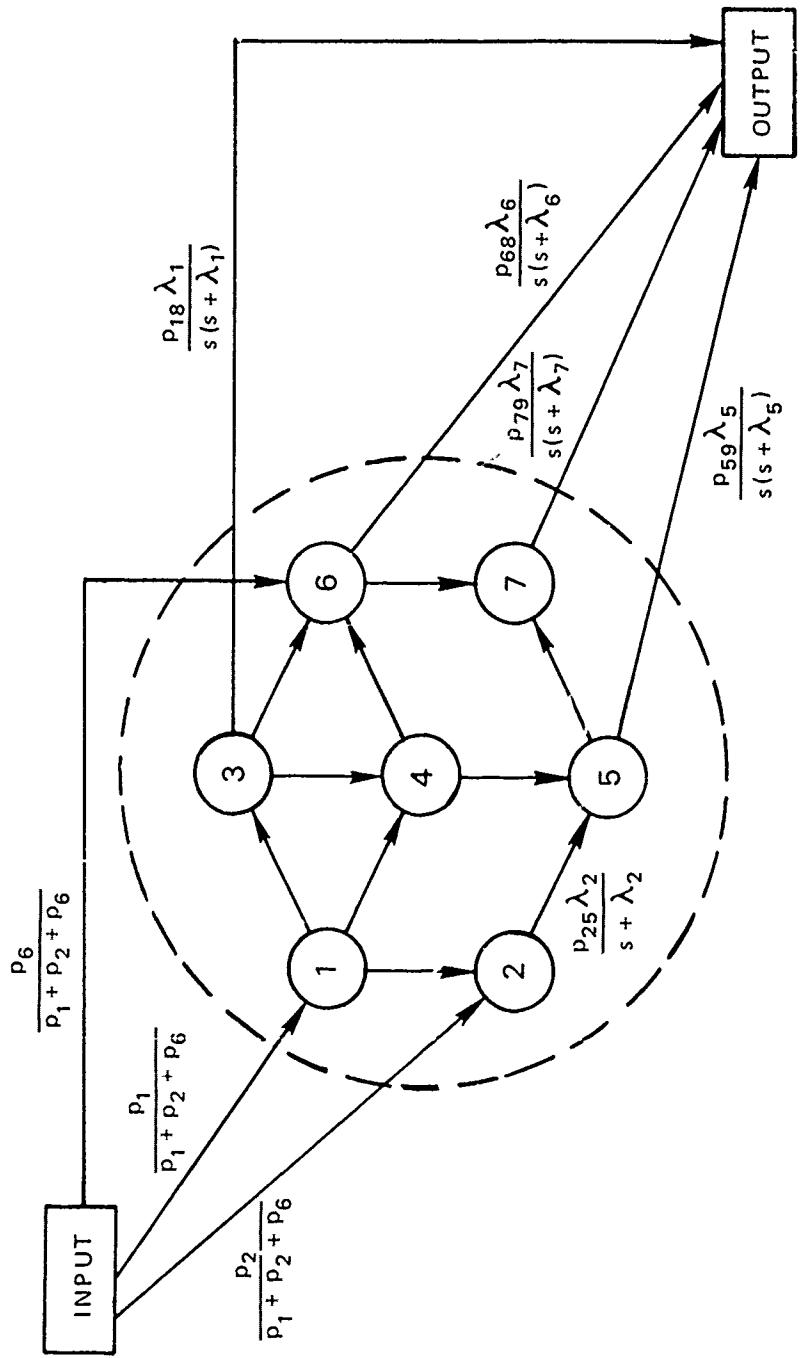


FIGURE A-9a SEMI-MARKOV FLOW GRAPH FOR TIME SPENT IN A SUBSET S OF STATES

the corresponding transition probabilities. Now add on all incoming arcs to S from the complementary set of states \bar{S} and join their tails at a fictitious source node labeled "INPUT," but, instead of having the original Laplace transforms associated with these arcs, designate them with the (normalized) steady-state probabilities of the states to which they lead.* Next, replace all outgoing arcs from set S by arcs to a fictitious node labeled "OUTPUT," and, in this case, multiply the original Laplace transforms on these arcs by the term $1/s$ (see Figure A-9a). It will not be demonstrated that the desired cumulative probability distribution is obtained as the inverse Laplace transform of the transmission of this flow graph from its input to its output.

Consider what happens when the system has just entered one of the states of the subset S from the complementary set of states \bar{S} . Given no further information, it can only be said that the (discrete) probability density for starting out consists of just the steady state (equilibrium) probabilities for the states entered from elsewhere in the flow graph only, that is, those states which have a non-zero transition probability to them. Once the system is in one of the states of S , it spends an amount of time t or less in S if, and only if, it reaches some state in \bar{S} by time t (for every $t > 0$). Thus, the desired cumulative probability distribution is clearly the sum over all possible initial states i of S and all possible adjacent states j in \bar{S} , of the terms:

$$P(\text{system is in state } i \text{ of } S \text{ and takes time } t \text{ to reach state } j | \text{system begins its stay in state } i) \cdot P(\text{system starts out in state } x_i) \quad (\text{A-24})$$

* The reason we do not take arcs from the "INPUT" node to all nodes of S is that we are interested only in those states of S at which we could have just arrived from the complement \bar{S} , and these are not necessarily all nodes of S .

Now, the transmission of a semi-Markov flow graph from a node i to a node j is, by definition the quantity:

$$\Psi_{ij}^* = \text{Lap. transf. of } P(\text{system is in state } j \text{ at time } t \mid \text{in } i \text{ at } t=0). \quad (\text{A-25})$$

Consider a modified flow graph of Figure A-9a, given in Figure A-9b, containing direct representations of the adjacent states j in S which are exited into upon leaving S . The interpretation of this new flow graph is that we now have expanded S to include its adjacent exit states which are now made to be so-called "trapping" states⁶ by the omission of all arcs emanating from them (including those to each other). Thus, when the transmission through the flow graph of Figure A-9b is computed, the Laplace transform of the probability of reaching any one of these trapping states by time t is obtained, i.e., the above sum of terms. Since this is the sum of transmissions to the trapping state nodes multiplied by $1/s$, Figures A-9a and A-9b are equivalent. Note that the output arcs in Figure A-9b have the factor $1/s$ only, not $1/(s + \lambda_j)$, since the transition rates λ_j for these states have been set equal to zero. Consequently, the above sum is just the inverse Laplace transform of the transmission from input to output for either equivalent flow graph of Figures A-9a or A-9b. This result is used in Sections IV-C and IV-D to represent and calculate saturated, unsaturated, and waiting time distributions for the single element ASW model. Note that the probability density function, instead of the cumulative function, is obtained by omitting the $1/s$ factors on the output arcs.

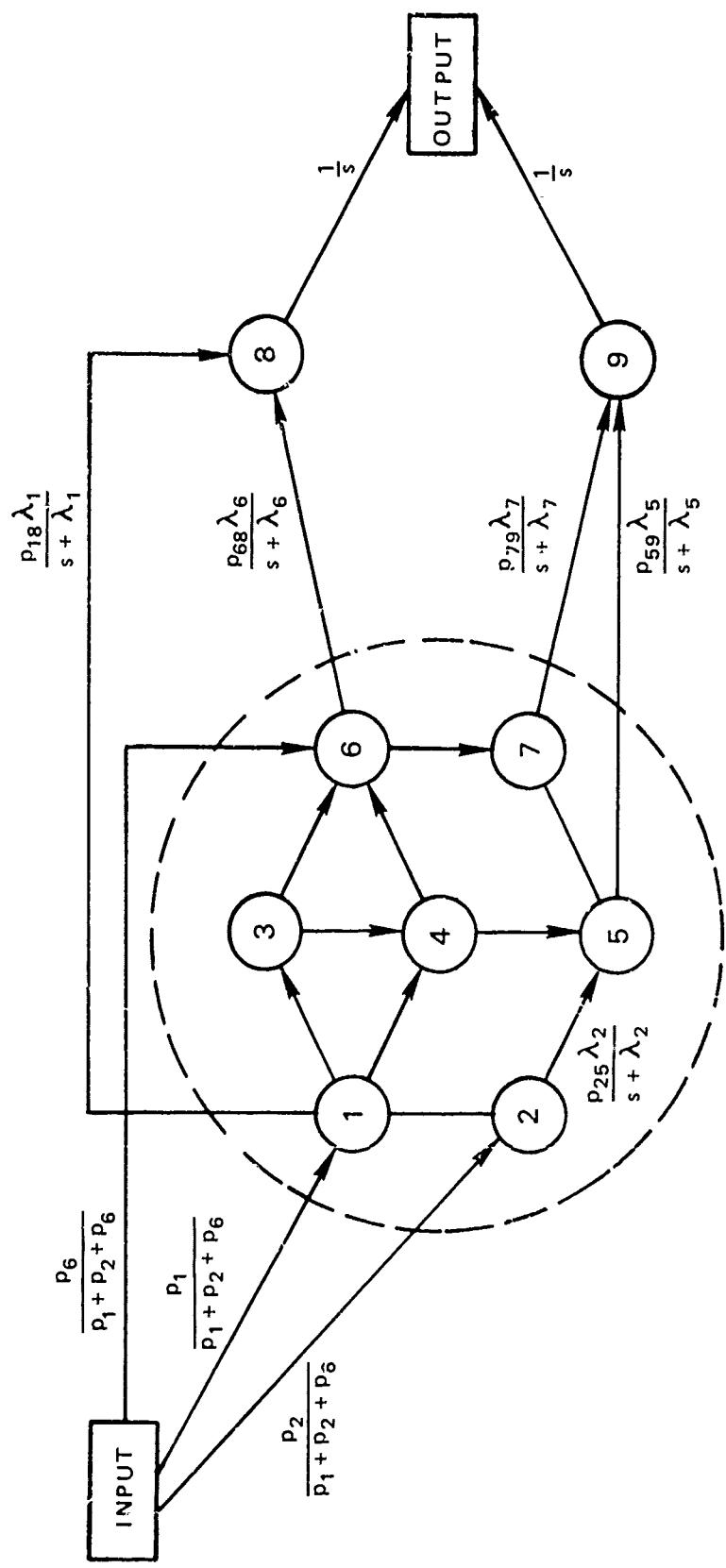


FIGURE A-9b EQUIVALENT FLOW GRAPH FOR TIME SPENT IN A SUBSET S OF STATES

Appendix B

**SOLVING THE SYSTEM PROBABILITY EQUATIONS
IN THE TRANSIENT STATE**

Appendix B

SOLVING THE SYSTEM PROBABILITY EQUATIONS IN THE TRANSIENT STATE

In this appendix a suitable system of differential equations for calculations of the desired cumulative probability equations, as functions of time, is developed. The general idea behind the "successive overrelaxation iterative method,"¹⁰ employed to compute solutions to these equations, is explained and this explanation is followed by a description of how graphtheoretical considerations led to great reduction in computing storage and number of operations required. The description will be with reference to Figure III-13, but the same overall approach was used for the other measures of effectiveness.

1. The Differential Equations

As in Appendix A, the system of differential equations to be solved in finding the probability distribution for the time t that a Markov process takes to reach a set of trapping states S_o is represented by the matrix differential equation (A-14)

$$\dot{\phi}(t) = \phi(t)A$$

where $A = [a_{ij}]$ is the transition rate matrix of the appropriate Markov process and is formed from a transition probability matrix P and a holding rate (diagonal) matrix Λ through the formula $A = \Lambda(P - I)$. These must also be associated with Figure III-13, but not the entire system, when considering system unsaturated time. These matrices in turn are obtainable from "competing process" holding rates λ_{ij} , which in this case arise

from the (assumed) independent "servers" comprising the queuing system, and are always equal to one of the quantities λ , u_1 , \dots , u_4 , or multiples thereof.

The desired probability distribution for unsaturated time is the quantity

$$\sum_{i \in S_I} \sum_{j \in S_0} p_i \varphi_{ij}(t) / \sum_{i \in S_I} p_i \quad (B-1)$$

where S_I and S_0 are the appropriate sets of input and output (absorbing) states, p_i is the steady state probability of state i , defined by Equation (A-9) in Appendix A, and φ_{ij} is defined by (A-11).

With reference to the general discussion of Appendix A, the appropriate input states for calculating system unsaturated time are the states with exactly 3 contacts in the system. That is, when the system enters an unsaturated condition, the number of contacts must have just changed from 4 to 3. Lacking any a priori information, the probability that any one of the substates in Figure III-7 is the one the system just entered since becoming unsaturated is just the steady state probability of that state divided by the sum of steady state probabilities for the 3 in-the-system states. Similar considerations lead to the designation of the appropriate output states for calculating system unsaturated time as the states with exactly 4 contacts in the system.

The quantity (B-1) is obtainable by starting off the integration of the system of differential equations

$$\dot{x}_i = \sum_{j=i} a_{ji} x_j \quad i = 1, 2, \dots, N \quad (B-2)$$

with the initial conditions

$$x_i(0) = \begin{cases} p_i / \sum_{i \in S_I} p_i & , \quad i \in S_I \\ 0 & , \quad i \notin S_I \end{cases} \quad (B-3)$$

and summing up the output state values to form:

$$x_0(t) = \sum_{j \in S_0} x_j(t) . \quad (B-4)$$

Equations (B-2) are the expanded form of the matrix differential equation (A-14), i.e., $\dot{\varphi} = \varphi A$, which has φ interpreted as a row vector to maintain conformity with Reference 6. To relate the following developments to Reference 10, it is necessary to realize that here we have the transpose of every matrix and vector employed in Reference 10.

The matrix A may be decomposed into lower and upper triangular and diagonal portions, because of the particular flow-graph structure of the model which allows grouping of nodes with 1, 2, 3, ..., etc., contacts in the

system. Furthermore, this decomposition of the matrix A is in a block diagonal form, having three "diagonals" of blocks of matrices, namely those corresponding to:

- i arrivals
- ii exits from the system, and
- iii completions of services at the interim queues within the system.

This structure is depicted in Figure B-1, for the case of unsaturated system time calculation.

Referring to Appendix A, the entries a_{ij} of the matrix A are just the appropriate "competing process" parameters λ_{ij} off the diagonal, and minus the sum of all off-diagonals λ_{ij} 's on the diagonal. In our model, it will be recalled that these λ_{ij} 's are always λ or multiples of one of the u_i 's, ($i=1$ to 4). The actual details of the block matrix A₂₂ in Figure B-1 are given in Figure B-2. The diagonal terms are not included for the sake of simplicity, as they are rather long to write, but the very first one, for example, would be $-\lambda - 2u_1 - u_4$.

$$A = \begin{bmatrix} A_{00} & A_{01} & 0 & 0 & 0 \\ A_{10} & A_{11} & A_{12} & 0 & 0 \\ 0 & A_{21} & A_{22} & A_{23} & 0 \\ 0 & 0 & A_{32} & A_{33} & A_{34} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \begin{cases} \text{None in the system} \\ \} 1 \quad " \quad " \quad " \\ \} 2 \quad " \quad " \quad " \\ \} 3 \quad " \quad " \quad " \\ \} \text{Corresponds to trapping states,} \\ \text{i.e., representatives of 4 in} \\ \text{the system} \end{cases}$$

FIGURE B-1 BLOCK DIAGONAL STRUCTURE OF THE ONE MATRIX A

$$A_{22} = \begin{bmatrix} \bullet & 2\mu_1 & 0 & 0 & & \\ & \bullet & \mu_2 & 0 & \mu_1 & \\ & & \bullet & \mu_3 & 0 & \mu_1 \\ & & & \bullet & 0 & 0 & \mu_1 \\ & A_{22} = & & & \bullet & 2\mu_2 & 0 & 0 & \\ & & & & & \bullet & \mu_3 & \mu_2 & 0 & \\ & & & & & & \bullet & 0 & \mu_2 & 0 & \\ & & & & & & & \bullet & 2\mu_3 & 0 & \\ & & & & & & & & \bullet & \mu_3 & \\ & & & & & & & & & \bullet & \\ & & & & & & & & & & \bullet \end{bmatrix}$$

FIGURE B-2 DETAILS OF ONE BLOCK MATRIX IN FIGURE B-1

2. The Successive Overrelaxation Iterative Method¹⁰

The following procedure is a convergent scheme for solving a system of linear equations of the form

$$vQ = g \quad (B-5)$$

where v, g are row vectors

$$v = \{v_1, \dots, v_N\}, \quad g = \{g_1, \dots, g_N\} \quad (B-6)$$

and Q is an $N \times N$ matrix which satisfies certain properties set out in Reference 10. This is not a priori an integration method, but it is found in integration methods that require subroutines involving solution of linear equations.

Let the matrix Q be written as a sum of upper triangular, diagonal,
and lower triangular matrices*

$$Q = D + E + F \quad (B-7)$$

where D is diagonal, E is upper triangular, and F is lower triangular. In fact, the very same general approach will work for block partitioned matrices Q (as we have in this case), for which D is taken as a diagonal of blocks instead of scalars. In the latter case, Reference 10 refers to the method as the block successive overrelaxation iterative method.

Now, suppose that a trial solution vector $v^{(k)}$ is available. The next trial solution vector $v^{(k+1)}$ is defined for this method by

$$v^{(k+1)}(D + \omega E) + v^{(k)} [\omega F - (1 - \omega)D] = \omega g \quad (B-8a)$$

where ω lies in the range $0 < \omega < 2$, and is called a relaxation factor.

The above relation is not as transparent as its following equivalent:

$$[v^{(k+1)} - v^{(k)}] [D + \omega E] = \omega [g - v^{(k)} Q] \quad (B-8b)$$

* As stated earlier, this discussion differs somewhat from Reference 10 in notation and conventions.

In other words, the iteration takes the error vector, if any, resulting from trying to solve Equation (B-5) and causes a particular

$$\Delta v^{(k)} = v^{(k+1)} - v^{(k)} \quad (B-9)$$

to reduce that error. When $\omega = 1$, the method reduces to the so-called Jacobi iterative method. When $\omega > 1$, there is what is called overrelaxation, i.e., "jumping ahead of yourself," which greatly reduces computation time. Reference 10 develops a formula for the optimal relaxation factor, which depends on the largest eigenvalue of a certain associated matrix. In our computations, it was found that the fixed value of $\omega = 1.3$ worked satisfactorily, but in future work with larger flow graphs arising from practical problems it may pay to determine an optimal ω .

It will be observed that Equation (B-8) does not give $v^{(k+1)}$ explicitly in terms of $v^{(k)}$ and fixed data. However, the fact that the matrix $D + \omega E$ is upper triangular allows for successive calculation of $v_1^{(k+1)}$, then $v_2^{(k+1)}$ in terms of $v_1^{(k+1)}$, etc., until $v_N^{(k+1)}$ is calculated in terms of the earlier $v_i^{(k+1)}$'s.

Now we come to the actual integration steps. As implied earlier, successive overrelaxation steps were used to solve linear equations within an integration step. Rather than use any standard available numerical integration routine, it was decided to take advantage of the sparsity of matrices associated with this model by preparing a special purpose computer program. In fact, the size of even the small-scale prototype model would preclude the use of a standard general purpose routine. The integration steps used were of the form:

$$\varphi(t + \Delta t) = \varphi(t) e^{A\Delta t} . \quad (B-10)$$

In view of the particular sparse structure and block diagonal form of the model, the use of a Padé approximation to the exponential of the matrix $A\Delta t$, in the form:

$$e^{A\Delta t} \approx \left(I - \frac{\Delta t}{2} A \right)^{-1} \left(I + \frac{\Delta t}{2} A \right) \quad (B-11)$$

or better yet,

$$e^{A\Delta t} \approx \left(I - \frac{\Delta t}{2} A + \frac{(\Delta t)^2}{12} A^2 \right)^{-1} \left(I + \frac{\Delta t}{2} A + \frac{(\Delta t)^2}{12} A^2 \right) \quad (B-12)$$

was recommended. Higher order approximations were considered and prepared for coding but were not actually used. Their use may be of value in larger practical problems, as well as in completion of calculations for the last measure of effectiveness, i.e., the contact time-in-the-system, which was not completed in time for inclusion in this report. The exponential of a matrix, incidentally, arise in this context simply because the formal solution to Equation (A-14) is $\varphi = \varphi_0 e^{At}$. Numerically, the situation is not that simple, however. A large matrix A is never actually exponentiated, e.g., with the Taylor series, since convergence of most series is very slow, especially for certain values of Δt . As an example, $e^{A\Delta t}$, even though it may have all relatively large negative real parts for its eigenvalues, would take incredibly long to calculate for large values of Δt . The reader may wish to verify this for himself by examining the series:

$$e^{-100} = 1 - 100 + 100^2/2 - \dots$$

Any approximation of $e^{A\Delta t}$ would have to be used for appropriately confined ranges of Δt . In this case, Equation (B-12) permitted much larger

increments of Δt than would be allowed using the simple approximation (used in Euler integration):

$$e^{A\Delta t} \approx I + A\Delta t .$$

Equation (B-5) arose as a result of the Padé approximation approach to integration in the following manner. Taking Equation (B-11) as a basis, and letting t be some time in the process of integration, we consider:

$$g = \varphi(t) \left[I + \frac{\Delta t}{2} A \right] \quad (B-13)$$

$$Q = I - \frac{\Delta t}{2} A \quad (B-14)$$

$$v = \varphi(t + \Delta t) .$$

To obtain $\varphi(t + \Delta t)$ from $\varphi(t)$, Equation (B-5) is solved for v with the above forms for Q and g . Setting $v^{(0)} = \varphi(t)$ is a good starting point for the iteration. For the illustrative computations, both Equation (B-11) and Equation (B-12) were used, and it required only about 4 to 10 iterations to achieve an error of less than 10^{-5} in the trial updated probabilities $v_j^{(k+1)}$. The only effect of going to Equation (B-12) was that use of larger Δt 's was possible without a reduction in accuracy than when using Equation (B-11).

3. Using the Flow Graph Structure to Organize Calculations Efficiently

Because of the fact that the matrix A for this particular problem is very sparse, all data are stored in terms of arc-node incidence relations for the appropriate flow graph, depending upon which measure of effectiveness is being quantified. In the case of system unsaturated time, the

sparsity was actually of the smallest degree since this case led to a total of only $N = 62$ states.

The number of nonzero a_{ji} 's for a given i in Equation (B-2) is anywhere from 1 to 5. Actually, any number of a_{ji} 's corresponding to incoming arcs might have been non-zero in a general flow graph, even though it has a limited number of outgoing arcs at each node, but the particular periodic sort of structure of the single element model's flow graphs assures no more than 5 arcs coming into any node as well as out of any node except for output nodes. Because of this special structure of all of the flow graphs, no matrix algebra was performed on the computer. Rather, we stored and worked with lists of arc factors, doubly indexed arrays, e.g., FACTOR(I,J), in which the first index ran from 1 to N (the total number of nodes) and J ran from 1 to 5. This led to a great saving in computing storage and effort as compared to using a general purpose integration scheme that would require manipulation of $N \times N$ matrices.

DISTRIBUTION LIST

Organization	No. of Copies	Organization	No. of Copies
Assistant Secretary of Defense (System Analysis) Washington, D.C. 20350	3	Commander Submarine Development Group TWO Naval Submarine Base New London Box 70 Groton, Connecticut 06340	1
Joint Chiefs of Staff Studies Analysis and Gaming Agency Washington, D.C. 20301	1	Naval War College Newport, Rhode Island 02840 (Technical Library)	1
Advanced Research Projects Agency Department of Defense Washington, D.C. 20301 (Technical Library)	1	Professor J. R. Borsting Department of Operations Research and Administrative Sciences (Code 55) Naval Postgraduate School Monterey, California 93940	1
Defense Documentation Center Cameron Station Alexandria, Virginia 22314	12	Professor J. D. Esary Department of Operations Research and Administrative Sciences Naval Postgraduate School Monterey, California 93940	1
Chief of Naval Operations Department of the Navy Washington, D.C. 20350 (Op-95) (Op-96) (Op-326E)	1 1 1	Professor Peter A. W. Lewis Department of Operations Research and Administrative Sciences Naval Postgraduate School Monterey, California 93940	1
Office of Naval Research Arlington, Virginia 22217 (Code 462) (Code 463) (Code 466) (Code 432) (Code 434)	2 1 1 1 1	Professor Peter C. C. Wang Department of Mathematics Naval Postgraduate School Monterey, California 93940	1
Commander, Submarine Force U.S. Atlantic Fleet Norfolk, Virginia 23511 (CDR Roberts)	1	Naval Postgraduate School Monterey, California 93940 (Technical Library)	1
Commander, Antisubmarine Warfare Force U.S. Pacific Fleet FPO San Francisco 96610 (Project UPTIDE Officer)	1	Naval Academy Annapolis, Maryland 21402 (Technical Library)	1
Commander, Antisubmarine Warfare Force U.S. SIXTH Fleet FPO New York 09521 Mr. Russell Coile)	1	Dr. William J. Kejka (Code 4300) Naval Electronics Laboratory Center Bldg. 33, Room 2045A San Diego, California 92152	1

DISTRIBUTION LIST (Continued)

<u>Organization</u>	<u>No. of Copies</u>	<u>Organization</u>	<u>No. of Copies</u>
Naval Electronics Laboratory Center Building 33, Room 2045A San Diego, California 92152 (Technical Library)	1	Director NATO SACLANT ASW Research Centre FPO New York 09019 (ATTN: Library)	1
Naval Ship Research and Development Center Bethesda, Maryland 20034	1	Professor Charles E. Antoniak Department of Statistics University of California Berkeley, California 94720	1
Naval Undersea Research and Development Center San Diego, California 92132 (Technical Library)	1	Center for Naval Analyses 1401 Wilson Boulevard Arlington, Virginia 22209 (Library)	1
Naval Undersea Research and Development Center Pasadena Laboratory 3202 East Foothill Boulevard Pasadena, California 91107	1	Professor K. T. Wallenius Department of Mathematics Clemson University Clemson, South Carolina 29631	1
Navy Underwater Systems Center Newport Laboratory Newport, Rhode Island 02840	2	Professor William F. Lucas Operations Research Department Cornell University Ithaca, New York 14850	1
Navy Underwater Systems Center New London Laboratory New London, Connecticut 06320	2	General Dynamics Electric Boat Division Groton, Connecticut 06340 (Dr. R. S. Hall, Code 413)	1
Naval Weapons Center China Lake, California 93555 (Library) (Code 12)	1	HHRB-Singer Company P. O. Box 60 State College, Pennsylvania 16801 (Dr. D. Smith)	1
Naval Research Laboratory Washington, D.C. 20390 (Technical Information Division) (Code 2629)	6	Applied Physics Laboratory Johns Hopkins University 8621 Georgia Avenue Silver Spring, Maryland 20901	1
Naval Ordnance Laboratory Silver Spring, Maryland 20910 (Library) (ASW-12 E. A. Letow, SAO, MASW Proj Office)	2	Institute of Defense Analyses 400 Army-Navy Drive Arlington, Virginia 22202	1
Dr. M. A. Thomas Head, Math Statistics Branch Naval Weapons Laboratory Dahlgren, Virginia 22448	1	Professor Ralph L. Disney Department of Industrial Engineering University of Michigan Ann Arbor, Michigan 48104	1

DISTRIBUTION LIST (Concluded)

<u>Organization</u>	<u>No. of Copies</u>
The Ohio State University 1775 South College Road 308 Hagerty Hall Columbus, Ohio 43210 (Dr. D. Howland)	1
Operations Research, Inc. 1400 Spring Street Silver Spring, Maryland 20910	1
Dr. Walter S. Liggett, Jr. Submarine Signal Division Raytheon Company Portsmouth, Rhode Island 02871	1
Professor John Walsh Department of Statistics Southern Methodist University Dallas, Texas 75222	1
Professor Frederick S. Hillier Department of Operations Research Stanford University Stanford, California 94305	1
Professor H. O. Hartley Director, Institute of Statistics Texas A&M University College Station, Texas 77840	1
Vector Research, Inc. P.O. Box 1506 Ann Arbor, Michigan 48016	1
Daniel H. Wagner Associates, Inc. Station Square One Paoli, Pennsylvania 19301	1
Professor Martin Shubik Department of Economics Yale University New Haven, Connecticut 06520	1

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Stanford Research Institute 333 Ravenswood Avenue Menlo Park, California 94025	2a. REPORT SECURITY CLASSIFICATION Unclassified
	2b. GROUP N/A

3. REPORT TITLE

DYNAMIC ANALYSIS OF ASW EFFECTIVENESS - A QUEUING APPROACH

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Techn. cal Report 1

5. AUTHOR(S) (First name, middle initial, last name)

Marvin W. Zumwalt Andrew J. Korsak Robert S. Ratner

6. REPORT DATE March 1972	7a. TOTAL NO. OF PAGES 182	7b. NO. OF REFS 17
8a. CONTRACT OR GRANT NO. N00014-71-C-0417	9a. ORIGINATOR'S REPORT NUMBER(S)	
b. PROJECT NO. RF 018-96-01	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	
c. NR 274-003-37		
d.		
10. DISTRIBUTION STATEMENT "Distribution limited to U.S. Government agencies only; test and evaluation: 9 August 1972. Other requests for this document must be referred to Office of Naval Research (Code 462)."		
11. SUPPLEMENTARY NOTES Reproduction in whole or in part is permitted for any purpose of the United States Government	12. SPONSORING MILITARY ACTIVITY Naval Analysis Programs (Code 462) Office of Naval Research Arlington, Virginia 22209	

13. ABSTRACT

This interim report demonstrates the feasibility of conducting a dynamic analysis of ASW effectiveness through the application of queuing theory. The ASW effectiveness of both the individual ASW unit and the aggregated ASW force are considered. Analytical models are formulated representing both levels of capability, and appropriate computational techniques are described and illustrated by selected examples. Areas requiring further research and development are defined, and recommendations for near term research are included which are in consonance with a continuing ASW evaluation program.

UNCLASSIFIED
Security Classification

KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Antisubmarine warfare						
Attack						
Detection						
Effectiveness methodology						
Flow graph analysis						
Localization						
Markov processes						
Model						
Queueing						
Semi-Markov processes						
Symbolic algebra						
System effectiveness						